

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:44:25 ; Search time 43.583 Seconds
(without alignments)
2301.458 Million cell updates/sec

Title: US-09-886-349A-2

Perfect score: 1802

Sequence: 1 MSNSRRSLRWLSVLA.....QTKSGTRGTVTLAEGPPA 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003ss:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1802	100.0	355	2	Aaw32367 Mycobacte
2	1802	100.0	355	2	Aaw32435 Mycobacte
3	1802	100.0	355	2	Aaw64307 Mycobacte
4	1802	100.0	355	2	Aaw81670 M. tuberc
5	1802	100.0	355	2	Aay38972 M. tuberc
6	1802	100.0	355	2	Aay39109 M. tuberc
7	1802	100.0	355	4	Aau01890 M. tuberc
8	1802	100.0	355	5	Aae29701 Mycobacte
9	1802	100.0	355	5	Aae17565 Mycobacte
10	1792	99.4	355	2	Aay05000 Mycobacte
11	1792	99.4	355	4	Aao22137 Mycobacte
12	1792	99.4	355	4	Aag81110 Mycobacte
13	1792	99.4	379	2	Aay04830 Mycobacte
14	1634	90.7	330	5	Aae29702 Mycobacte
15	1634	90.7	330	5	Aae17566 Mycobacte
16	1631	90.5	330	5	Aae29703 Mycobacte
17	1631	90.5	330	5	Aae17567 Mycobacte
18	1631	90.5	330	7	ADA26372 Mycobacte
19	1631	90.5	723	7	ADA26354 Mycobacte
20	1631	90.5	1010	7	ADA26356 Mycobacte
21	1627	90.3	330	7	ADA26371 Mycobacte
22	999.5	55.5	1016	7	ADA26370 M. bovis
23	992	55.0	1154	7	ADA26368 Mycobacte
24	990.5	55.0	825	7	ADA26366 Mycobacte
25	989.5	54.9	813	7	ADA26367 Mycobacte

26	988	54.8	1022	7	ADA26369	Mycobacte
27	985	54.7	195	5	Aae29704	Mycobacte
28	985	54.7	195	5	Aae17568	Mycobacte
29	983.5	54.6	875	7	ADA26365	Mycobacte
30	983	54.6	596	2	Aay32070	Mycobacte
31	983	54.6	596	5	Aae29710	Mycobacte
32	983	54.6	596	5	Aae17574	Mycobacte
33	983	54.6	599	5	Aau74599	Antigenic
34	983	54.6	729	4	Aao22142	Ral12-H9-3
35	983	54.6	729	5	Aae29708	Mycobacte
36	983	54.6	729	5	Aae17572	Mycobacte
37	983	54.6	729	7	ADA26373	Mycobacte
38	983	54.6	930	5	Aae29731	Mycobacte
39	983	54.6	930	7	ADA26364	Mycobacte
40	980	54.4	729	5	Aae29709	Mycobacte
41	980	54.4	729	5	Aae17573	Mycobacte
42	980	54.4	729	7	ADA26374	Mycobacte
43	978	54.3	726	5	Aau74588	Antigenic
44	978	54.3	729	2	Aay32059	Mycobacte
45	676	37.5	132	2	Aaw32354	Mycobacte

ALIGNMENTS

RESULT 1
AAW32367

ID AAW32367 standard; protein; 355 AA.

AC AAW32367;

XX

DT 13-JAN-1998 (first entry)

XX

DE Mycobacterium tuberculosis antigen TBra35.

XX

KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

KW skin testing; M.tuberculosis.

XX

OS Mycobacterium tuberculosis.

XX

PN WC9709429-A2.

XX

PD 13-MAR-1997.

XX

PF 30-AUG-1996; 96WO-US014675.

XX

PR 01-SEP-1995; 95US-00523435.

PR

PR 22-SEP-1995; 95US-00532136.

PR

PR 22-MAR-1996; 96US-00620280.

PR

PR 05-JUN-1996; 96US-00658800.

PR

PR 12-JUL-1996; 96US-00680573.

XX

(CORI-) CORIXA CORP.

PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;

PI Vedvick TH, Twardzik DR;

XX

DR WPI; 1997-192904/17.

XX

DR N-PSDB; AAT91414.

XX

PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -

XX

PS useful for diagnosis of M. tuberculosis infection.

XX

PS Example 3; Page 124-126; 190pp; English.

XX

CC A new immunogenic polypeptide has been developed comprising an

CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its

CC variant differing only in conservative substitutions and/or

CC modifications). The present sequence represents a M.tuberculosis antigen,

CC TBra35. The immunogenic polypeptide can be used to diagnose

CC M.tuberculosis infection by forming complexes with specific antibodies in

CC the sample. Fragments of DNA encoding the immunogenic polypeptide can be

CC used as diagnostic primers or probes and agents that bind to the antigen,

CC especially monoclonal antibodies or equivalent polyclonal antibodies, are
 CC also used for diagnosis

SQ Sequence 355 AA;
 Query Match 100.0%; Score 1802; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MNSRRSLRWSSLLSVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60
 Db 1 MNSRRSLRWSSLLSVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNVINTKLYNNVAGAGTGVIDPENGVLTTNNHVIAGATDINAFSVGSGQTYGVVDVVG 120
 Db 61 PQVNVINTKLYNNVAGAGTGVIDPENGVLTTNNHVIAGATDINAFSVGSGQTYGVVDVVG 120
 QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 Db 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 QY 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFOLSGQGGA 240
 Db 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFOLSGQGGA 240
 QY 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAAPASLGIST 300
 Db 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAAPASLGIST 300
 QY 301 GDVITAVDGAIPINSATAMADALNGHPGDVISVNMWTKSGGTRTGNVTLAEGPPA 355
 Db 301 GDVITAVDGAIPINSATAMADALNGHPGDVISVNMWTKSGGTRTGNVTLAEGPPA 355

RESULT 2
 AAW32435
 ID AAW32435 standard; protein; 355 AA.
 AC AAW32435;
 XX 08-JAN-1998 (first entry)
 DT Mycobacterium tuberculosis antigen TbrA35.
 DE Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 KW skin testing; M.tuberculosis.
 XX Mycobacterium tuberculosis.
 OS WO9709428-A2.
 PN 13-MAR-1997.
 PD 30-AUG-1996; 96WO-US014674.
 PF 01-SEP-1995; 95US-00523436.
 PR 22-SEP-1995; 95US-00533634.
 PR 22-MAR-1996; 96US-00620874.
 PR 05-JUN-1996; 96US-00659683.
 PR 12-JUL-1996; 96US-00680574.
 XX (CORI-) CORIXA CORP.
 PA Read SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TH, Twardzik DR;
 XX WPI; 1997-192903/17.
 DR N-PSDB; AAT91477.
 XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also for
 PT diagnosis.

PS Example 3; Page 114-116; 168pp; English.

XX A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
 CC variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis antigen,
 CC TbrA35. The immunogenic protein, and fusion proteins containing one or
 CC more of the proteins or one of the proteins plus ESAT-6, are useful in
 CC vaccines, preferably when formulated with a non-specific adjuvant, to
 CC induce an immune response against M.tuberculosis (for treatment or
 CC prevention)

SQ Sequence 355 AA;

Query Match 100.0%; Score 1802; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MNSRRSLRWSSLLSVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60
 Db 1 MNSRRSLRWSSLLSVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNVINTKLYNNVAGAGTGVIDPENGVLTTNNHVIAGATDINAFSVGSGQTYGVVDVVG 120
 Db 61 PQVNVINTKLYNNVAGAGTGVIDPENGVLTTNNHVIAGATDINAFSVGSGQTYGVVDVVG 120
 QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 Db 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 QY 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFOLSGQGGA 240
 Db 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFOLSGQGGA 240
 QY 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAAPASLGIST 300
 Db 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAAPASLGIST 300
 QY 301 GDVITAVDGAIPINSATAMADALNGHPGDVISVNMWTKSGGTRTGNVTLAEGPPA 355
 Db 301 GDVITAVDGAIPINSATAMADALNGHPGDVISVNMWTKSGGTRTGNVTLAEGPPA 355

RESULT 3
 AAW64307
 ID AAW64307 standard; protein; 355 AA.
 AC AAW64307;
 XX 17-OCT-2003 (revised)
 DT 09-NOV-1998 (first entry)

DE Mycobacterium tuberculosis antigen TbrA35.
 KW Tuberculosis; infection; diagnosis; antigen; TbrA35.
 XX Mycobacterium tuberculosis;
 OS WO9816645-A2.
 PN 23-APR-1998.
 PD 07-OCT-1997; 97WO-US018214.
 PF 11-OCT-1996; 96US-00729622.
 PR 13-MAR-1997; 97US-00818111.
 XX (CORI-) CORIXA CORP.
 PA Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX WPI; 1998-251292/22.

DR N-PSDB; AAV44355.

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to

PT develop products for the detection of M. tuberculosis infection and

PT diagnosis of tuberculosis.

XX

PS Example 3; Page 115-116; 250pp; English.

XX

CC This polypeptide comprises Mycobacterium tuberculosis soluble antigen

CC TBRa35. It is encoded by a DNA sequence (see AAV44355) isolated from a M.

CC tuberculosis strain H37Ra expression library with rabbit anti-sera raised

CC against M. tuberculosis supernatant. No significant homology was found

CC between TBRa35 and Genbank database sequences. The invention relates to

CC compositions and methods for diagnosing tuberculosis. It provides

CC polypeptides (see AAV44291-464379) comprising an antigenic portion of a

CC soluble M. tuberculosis antigen, or an immunogenic portion of an M.

CC tuberculosis antigen, as well as DNA sequences encoding such

CC polypeptides, recombinant expression vectors and transformed or

CC transfected host cells. Also claimed are methods and diagnostic kits for

CC detecting M. tuberculosis infection in a patient using the above

CC polypeptides, antibodies or oligonucleotide probes and primers, for the

CC diagnosis of tuberculosis. (Updated on 17-OCT-2003 to standardise OS

CC field)

XX

SQ Sequence 355 AA;

Query Match 100.0%; Score 1802; DB 2; Length 355;

Best Local Similarity 100.0%; Pred. No. 7.2e-126;

Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60

DB 1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60

QY 61 PQVNVNTKLYNNVAVGAGTGVIDPVGVLNNHVIAGATDINAFSGSGQTYGVDVVG 120

DB 61 PQVNVNTKLYNNVAVGAGTGVIDPVGVLNNHVIAGATDINAFSGSGQTYGVDVVG 120

QY 121 YDRTQDVAVLQIRGAGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180

DB 121 YDRTQDVAVLQIRGAGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180

QY 181 QASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFOLSQGGQGFA 240

DB 181 QASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFOLSQGGQGFA 240

QY 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAAPAASLGIST 300

DB 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAAPAASLGIST 300

QY 301 GDVITAVDGAIPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

DB 301 GDVITAVDGAIPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 4

AAW81670

ID AAW81670 standard; protein; 355 AA.

XX

AC AAW81670;

XX

DT 27-JAN-1999 (first entry)

XX

DE M. tuberculosis immunogenic polypeptide TBRa35.

XX

DE Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;

KW vaccine; pharmaceutical; infection; diagnosis.

XX

OS Mycobacterium tuberculosis.

XX

PN WO9816646-A2.

XX

PD 23-APR-1998.

XX 07-OCT-1997; 97WO-US018293.

XX

PR 11-OCT-1996; 96US-00730510.

PR 13-MAR-1997; 97US-00818112.

XX

PA (CORI-) CORIXA CORP.

XX

PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;

PI Vedwick TS, Twardzik DR, Lodes MJ;

XX

DR WPI; 1998-261042/23.

DR N-PSDB; AAV64463.

XX

PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to

PT develop products for the detection of M. tuberculosis infection and for

PT diagnosis, treatment and prevention of tuberculosis.

XX

PS Example 3; Page 110-111; 230pp; English.

XX

CC This sequence represents an immunogenic portion of a soluble

CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for

CC inducing protective immunity against tuberculosis (TB). This sequence can

CC be formulated into vaccines and/or pharmaceutical compositions for

CC immunising against M. tuberculosis infection or may be used for the

CC diagnosis of tuberculosis

XX

SQ Sequence 355 AA;

Query Match 100.0%; Score 1802; DB 2; Length 355;

Best Local Similarity 100.0%; Pred. No. 7.2e-126;

Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60

DB 1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60

QY 61 PQVNVNTKLYNNVAVGAGTGVIDPVGVLNNHVIAGATDINAFSGSGQTYGVDVVG 120

DB 61 PQVNVNTKLYNNVAVGAGTGVIDPVGVLNNHVIAGATDINAFSGSGQTYGVDVVG 120

QY 121 YDRTQDVAVLQIRGAGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180

DB 121 YDRTQDVAVLQIRGAGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180

QY 181 QASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFOLSQGGQGFA 240

DB 181 QASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFOLSQGGQGFA 240

QY 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAAPAASLGIST 300

DB 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAAPAASLGIST 300

QY 301 GDVITAVDGAIPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

DB 301 GDVITAVDGAIPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 5

AAV38972

ID AAV38972 standard; protein; 355 AA.

XX

AC AAV38972;

XX

DT 05-NOV-1999 (first entry)

XX

DE M. tuberculosis recombinant antigen protein TBRa35.

XX

KW Antigen; diagnosis; detection; infection; antibody; immunisation;

KW vaccine; immunity.

XX

OS Mycobacterium tuberculosis.

XX

PN	WO9942118-A2.	XX	Os	Mycobacterium tuberculosis.	XX
XX		XX			
PD	26-AUG-1999.	XX	PN	WO9942076-A2.	XX
PF		XX	PD		
PR	17-FEB-1999; 99WO-US003265.	XX	PF	26-AUG-1999.	XX
PR	18-FEB-1998; 98US-00024753.	XX	PR	17-FEB-1999; 99WO-US003268.	XX
PR	05-MAY-1998; 98US-00072596.	XX	PR	18-FEB-1998; 98US-00025197.	XX
XX		XX	PR	05-MAY-1998; 98US-00072967.	XX
PA	(CORI-) CORIXA CORP.	XX	PA	(CORI-) CORIXA CORP.	XX
XX		XX	XX		
PI	Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;	XX	PI	Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;	XX
PI	Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;	XX	PI	Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;	XX
XX		XX	XX		
DR	WPI; 1999-527416/44.	XX	XX		
DR	N-PSDB; AAZ19053.	XX	XX		
XX		XX	XX		
PT	New polypeptide comprising antigenic portions of M. tuberculosis.	XX	XX		
XX		XX	XX		
PS	Example 3; Page 151-153; 323pp; English.	XX	XX		
CC	This invention describes novel recombinant antigens and their encoding	XX	XX		
CC	nucleic acids derived from Mycobacterium tuberculosis. The novel	XX	XX		
CC	polypeptides are useful for detecting M. tuberculosis infection in a	XX	XX		
CC	biological sample by detecting antibodies which bind with the	XX	XX		
CC	polypeptides, and are useful as vaccines for immunizing against M.	XX	XX		
CC	tuberculosis infection. The new detection methods are needed as current	XX	XX		
CC	vaccination strategies do not provide 100% immunity	XX	XX		
XX		XX	XX		
SQ	Sequence 355 AA;	XX	SQ	Sequence 355 AA;	XX
Query Match 100.0%; Score 1802; DB 2; Length 355;					
Best Local Similarity 100.0%; Pred. No. 7.2e-126;					
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60	QY	1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60	QY	1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
Db	1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60	Db	1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60	Db	1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
QY	61 PQVNVNINIKLYNNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120	QY	61 PQVNVNINIKLYNNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120	QY	61 PQVNVNINIKLYNNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
Db	61 PQVNVNINIKLYNNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120	Db	61 PQVNVNINIKLYNNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120	Db	61 PQVNVNINIKLYNNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
QY	121 YDRTDQVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGGCTPRAVPGRVVALLGQTV 180	QY	121 YDRTDQVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGGCTPRAVPGRVVALLGQTV 180	QY	121 YDRTDQVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGGCTPRAVPGRVVALLGQTV 180
Db	121 YDRTDQVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGGCTPRAVPGRVVALLGQTV 180	Db	121 YDRTDQVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGGCTPRAVPGRVVALLGQTV 180	Db	121 YDRTDQVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGGCTPRAVPGRVVALLGQTV 180
QY	181 QASDSLITGAEETLNGLIQFDDAAIQPGDSGGPVVNGLGQVVGMTAASDNFQLSQGGGGA 240	QY	181 QASDSLITGAEETLNGLIQFDDAAIQPGDSGGPVVNGLGQVVGMTAASDNFQLSQGGGGA 240	QY	181 QASDSLITGAEETLNGLIQFDDAAIQPGDSGGPVVNGLGQVVGMTAASDNFQLSQGGGGA 240
Db	181 QASDSLITGAEETLNGLIQFDDAAIQPGDSGGPVVNGLGQVVGMTAASDNFQLSQGGGGA 240	Db	181 QASDSLITGAEETLNGLIQFDDAAIQPGDSGGPVVNGLGQVVGMTAASDNFQLSQGGGGA 240	Db	181 QASDSLITGAEETLNGLIQFDDAAIQPGDSGGPVVNGLGQVVGMTAASDNFQLSQGGGGA 240
QY	241 IPIQQAIAIAGIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQRVVGSAAPASLGIST 300	QY	241 IPIQQAIAIAGIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQRVVGSAAPASLGIST 300	QY	241 IPIQQAIAIAGIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQRVVGSAAPASLGIST 300
Db	241 IPIQQAIAIAGIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQRVVGSAAPASLGIST 300	Db	241 IPIQQAIAIAGIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQRVVGSAAPASLGIST 300	Db	241 IPIQQAIAIAGIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQRVVGSAAPASLGIST 300
QY	301 GDVITAVDGPINSATAMADALNGHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355	QY	301 GDVITAVDGPINSATAMADALNGHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355	QY	301 GDVITAVDGPINSATAMADALNGHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355
Db	301 GDVITAVDGPINSATAMADALNGHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355	Db	301 GDVITAVDGPINSATAMADALNGHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355	Db	301 GDVITAVDGPINSATAMADALNGHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355
RESULT 6					
AAU01890					
ID	AAU01890 standard; protein; 355 AA.	ID	AAU01890 standard; protein; 355 AA.	ID	AAU01890 standard; protein; 355 AA.
XX		XX		XX	
AC	AAU01890;	AC	AAU01890;	AC	AAU01890;
XX		XX		XX	
DT	05-NOV-1999 (first entry)	DT	05-NOV-1999 (first entry)	DT	05-NOV-1999 (first entry)
XX		XX		XX	
DE	M. tuberculosis antigen TBra35 amino acid sequence.	DE	M. tuberculosis antigen TBra35 amino acid sequence.	DE	M. tuberculosis antigen TBra35 amino acid sequence.
XX		XX		XX	
KW	Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;	KW	Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;	KW	Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW	immunotherapy; diagnosis; immunisation; vaccine; infection;	KW	immunotherapy; diagnosis; immunisation; vaccine; infection;	KW	immunotherapy; diagnosis; immunisation; vaccine; infection;
KW	immune response; skin test.	KW	immune response; skin test.	KW	immune response; skin test.

```
XX 29-AUG-2001 (first entry)
XX M. tuberculosis antigen Tba35 (Mtb32A).
XX
XX Tba35; Mtb32A; antigen; vaccine; tuberculosis; AIDS;
XX acquired immunodeficiency disease.
XX
XX Mycobacterium tuberculosis.
XX
XX WO200124820-A1.
XX
XX 12-APR-2001.
XX
XX 10-OCT-2000; 2000WO-US028095.
XX
XX 07-OCT-1999; 99US-0158338P.
XX
XX 07-OCT-1999; 99US-0158425P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Houghton RL, Mcneill PD, Dillon DC, Lodes ML;
XX WPI; 2001-290576/30.
XX N-PSDB; AAS03781.
XX
XX Vaccinating against Mycobacteria infections in mammals using fusion
XX proteins comprising combinations of heterologous antigens.
XX
XX Disclosure; Page 153-154; 168pp; English.
XX
XX The sequence represents Mycobacterium tuberculosis Tba35 (also known as
XX Mtb32A), an M. tuberculosis antigen. Compositions comprising at least 2
XX heterologous antigens, as a fusion protein, and vectors expressing the
XX fusion proteins are used as vaccines to prophylactically immunise mammals
XX (especially humans) against infection by Mycobacteria. The compositions
XX contain at least 2 heterologous antigens that increase the serological
XX sensitivity of individuals infected with tuberculosis, a disease
XX frequently affecting patients with acquired immunodeficiency disease,
XX AIDS
XX
XX Sequence 355 AA;
XX
Query Match 100.0%; Score 1802; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 7.2e-126;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
Db 1 MSNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
QY 61 PQVNTNTKLYNNVAGAGTGIVIDPVGVLTTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
Db 61 PQVNTNTKLYNNVAGAGTGIVIDPVGVLTTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
QY 121 YDRTDQVAVLQLRGAGLSAAGGVAVGEPVAVMGNSGGQGTTPRVPGRVWALGQTV 180
Db 121 YDRTDQVAVLQLRGAGLSAAGGVAVGEPVAVMGNSGGQGTTPRVPGRVWALGQTV 180
QY 181 QASDSLTAAGETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAASDNFQLSQGGQGPA 240
Db 181 QASDSLTAAGETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAASDNFQLSQGGQGPA 240
QY 241 IFIQAWAIAGQIRSGGSGPTVHIGFTAGLGVVDNNGNGARVQRVWGSFAAASLGIST 300
Db 241 IFIQAWAIAGQIRSGGSGPTVHIGFTAGLGVVDNNGNGARVQRVWGSFAAASLGIST 300
QY 301 GDVITAVDGAIPNSATAMADALNGHPEPGDVISVNWQTKSGGTRTGNTVLAEGPPA 355
Db 301 GDVITAVDGAIPNSATAMADALNGHPEPGDVISVNWQTKSGGTRTGNTVLAEGPPA 355
```

RESULT 8

```
AAE29701
ID AAE29701 standard; protein; 355 AA.
XX
XX AAE29701;
XX
XX 27-JAN-2003 (first entry)
XX
XX Mycobacterium tuberculosis Mtb32A antigenic protein.
XX
XX Vaccine; immunity; diagnostic agent; gene therapy; Mtb32A antigen;
XX Ra35FL.
XX
XX Mycobacterium tuberculosis.
XX
XX WO200272792-A2.
XX
XX 19-SEP-2002.
XX
XX 13-MAR-2002; 2002WO-US008223.
XX
XX 13-MAR-2001; 2001US-0275837P.
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Brannon M, Guderian J;
XX
XX WPI; 2002-759844/82.
XX N-PSDB; AAD47076.
XX
XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
XX M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
XX against pathogenic microorganisms e.g. Leishmania and Mycobacterium
XX tuberculosis.
XX
XX Disclosure; Page 78-79; 155pp; English.
XX
XX The invention relates to a recombinant nucleic acid molecule encoding a
XX fusion polypeptide. The recombinant nucleic acid comprises a heterologous
XX polynucleotide sequence encoding an antigen or an antigenic fragment from
XX Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
XX polypeptide or its fragment. The Leishmania polynucleotide is selected
XX from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
XX are used in methods for eliciting immune response in mammals. They are
XX useful as vaccines to elicit protective immunity against pathogenic
XX microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
XX polypeptides are used for enhancing the expression of polynucleotides, as
XX in vivo diagnostic agents and for raising antibodies in a non-human
XX animal. The invention is used in gene therapy. The present sequence is M.
XX tuberculosis Mtb32A antigenic protein. Mtb32A is also referred to as
XX Ra35FL
XX
XX Sequence 355 AA;
XX
Query Match 100.0%; Score 1802; DB 5; Length 355;
Best Local Similarity 100.0%; Pred. No. 7.2e-126;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
Db 1 MSNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
QY 61 PQVNTNTKLYNNVAGAGTGIVIDPVGVLTTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
Db 61 PQVNTNTKLYNNVAGAGTGIVIDPVGVLTTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
QY 121 YDRTDQVAVLQLRGAGLSAAGGVAVGEPVAVMGNSGGQGTTPRVPGRVWALGQTV 180
Db 121 YDRTDQVAVLQLRGAGLSAAGGVAVGEPVAVMGNSGGQGTTPRVPGRVWALGQTV 180
QY 181 QASDSLTAAGETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAASDNFQLSQGGQGPA 240
Db 181 QASDSLTAAGETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAASDNFQLSQGGQGPA 240
```

QY 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVNDNGNGARVQVVGSAAPAASLGIST 300
 DB 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVNDNGNGARVQVVGSAAPAASLGIST 300

QY 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWOTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWOTKSGGTRTGNVTLAEGPPA 355

RESULT 9
 AAEL17565
 ID AAE17565 standard; protein; 355 AA.
 AC AAEL17565;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Mycobacterium species MTB32A (Ra35FL) protein.
 XX
 KW Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB32A; Ra32FL protein.
 XX
 XX Mycobacterium sp.
 OS
 XX W0200198460-A2.
 XX
 XX 27-DEC-2001.
 XX
 XX 20-JUN-2001; 2001WO-US019959.
 PF
 XX 20-JUN-2000; 2000US-00597796.
 PR
 PR 01-FEB-2001; 2001US-0265737P.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX Skeiky Y, Reed S, Alderson M;
 PI
 XX WPI; 2002-147798/19.
 DR
 DR N-PSDB; AAD28335.
 XX
 XX Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 XX Claim 70; Page 94-95; 136pp; English.
 PS
 XX The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB32A (Ra32FL) protein
 XX
 XX Sequence 355 AA;

Query Match 100.0%; Score 1802; DB 5; Length 355;
 Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNRRRLRWSLLSLAAVGLGATAPAAQAPALSQDRFADFPALPLDPSAMVAQVA 60
 DB 1 MSNRRRLRWSLLSLAAVGLGATAPAAQAPALSQDRFADFPALPLDPSAMVAQVA 60

QY 61 PQVNVNTKLGNNAVGAGTGIVIDPNGVLTNNHVIAGATDINAFSVGSGQTVGVDDVVG 120
 DB 61 PQVNVNTKLGNNAVGAGTGIVIDPNGVLTNNHVIAGATDINAFSVGSGQTVGVDDVVG 120

QY 121 YDRTQDVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 DB 121 YDRTQDVAVLQIRGAGGLPSAAICGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180

QY 181 QASDLSLTGAETLNLQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSGGGGPPA 240
 DB 181 QASDLSLTGAETLNLQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSGGGGPPA 240

QY 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVNDNGNGARVQVVGSAAPAASLGIST 300
 DB 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVNDNGNGARVQVVGSAAPAASLGIST 300

QY 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWOTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWOTKSGGTRTGNVTLAEGPPA 355

RESULT 10
 AAY05000
 ID AAY05000 standard; protein; 355 AA.
 AC AAY05000;
 XX
 DT 06-JUL-1999 (first entry)
 XX
 DE Mycobacterium species protein sequence 50D.
 XX
 KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.
 XX
 XX Mycobacterium sp.
 OS
 XX W09909186-A2.
 FN
 XX 25-FEB-1999.
 PD
 XX 14-AUG-1998; 98WO-FR001813.
 PF
 XX 14-AUG-1997; 97FR-00010404.
 PR
 PR 11-SEP-1997; 97FR-00011325.
 XX
 XX (INSP) INST PASTEUR.
 XX
 XX Gicquel B, Portnoie D, Lim E, Pellicic V, Guigueno A;
 PI Goguet De La Salmoniere Y;
 XX
 XX WPI; 1999-181045/15.
 DR N-PSDB; AAX34251.
 XX
 XX Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in infection-associated
 PT protein expression.
 XX
 XX Claim 32; Fig 50D; 309pp; French.
 PS
 XX Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
 CC from various Mycobacterium species microorganisms. The encoding
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection

Query Match 99.4%; Score 1792; DB 2; Length 355;
 XX Sequence 355 AA;

[illegible]

RESULT 11	
AAO22137	
ID	AAO22137 standard; protein; 355 AA.
XX	
XX	
AC	AAO22137;
XX	
DT	03-OCT-2002 (first entry)
XX	
DE	Mycobacterium tuberculosis MTB32A protein.
XX	
KW	Ra12; serine protease antigen; MTB32A; Mycobacterium tuberculosis;
KW	vaccine; eukaryotic; prokaryotic origin; serum antibody; fusion protein;
KW	immunogen; cytokine.

CC	antibodies to <i>M. tuberculosis</i> antigens in an individual indicates that
CC	the individual is infected with it. The fusion polypeptides are useful as
CC	sources of proteins for monitoring binding of serum antibodies to fusion
CC	proteins and as an immunogen to induce and/or enhance immune responses.
CC	The coding sequences can be ligated with a coding sequence of another
CC	molecule such as <i>M. tuberculosis</i> antigens, a cytokine or an adjuvant and
CC	can be used in vivo as a DNA vaccine. This sequence represents the
CC	<i>Mycobacterium tuberculosis</i> MTE32A protein
XX	
XX	Sequence 355 AA;
XX	
XX	Query Match 99.4%; Score 1792; DB 4; Length 355;
XX	Best Local Similarity 99.4%; Pred. No. 4e-125;
XX	Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	1 MSNSRRSLRWSLLSVLAAVGLGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQYA 60
DB	1 MSNSRRSLRWSLLSVLAAVGLGLATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVG 60
QY	61 PQVYNINTKLGYNNAVAGTGIVIDPNGVLTNNHVIAGATDINAFSVGSGQTGYDVVVG 120
DB	61 PQVYNINTKLGYNNAVAGTGIVIDPNGVLTNNHVIAGATDINAFSVGSGQTGYDVVVG 120
QY	121 YDRQDVAVLQLRGAGGLPSAAIGGGVAVGPPVAMGNSGGGTPPRAVPGRVVVG 180
DB	121 YDRQDVAVLQLRGAGGLPSAAIGGGVAVGPPVAMGNSGGGTPPRAVPGRVVVG 180
QY	181 QASDSLTCGAETLNGLIQFDAAIQPGSDGGPVVNGLGQVWGMNTAASDNFQLSGQGQGPPA 240
DB	181 QASDSLTCGAETLNGLIQFDAAIQPGSDGGPVVNGLGQVWGMNTAASDNFQLSGQGQGPPA 240
QY	241 IPIQCAMAIAGIIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVORVVVGSAPAAASLGIST 300
DB	241 IPIQCAMAIAGIIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVORVVVGSAPAAASLGIST 300
QY	301 GDVITAVDGPINSATAMADALNGHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355
DB	301 GDVITAVDGPINSATAMADALNGHPGDVTSVNWQTKSGGTRTGNVTLAEGPPA 355
XX	
XX	RESULT 12
XX	AAAG81110
XX	ID AAAG81110 standard; protein; 355 AA.
XX	AC AAAG81110;
XX	DT 04-SEP-2001 (first entry)
XX	DE <i>Mycobacterium tuberculosis</i> potential drug target protein SEQ ID 161.
XX	DE <i>Mycobacterium tuberculosis</i> .
XX	DE Drug target; growth; organism viability; characterisation.
XX	DE <i>Mycobacterium tuberculosis</i> .

XX	WO200135317-A1.	
XX	EN	
XX	PD	17-MAY-2001.
XX	PF	13-NOV-2000; 2000WO-US031152.
XX	PR	12-NOV-1999; 99US-0165086P.
XX	PR	12-NOV-1999; 99US-0165124P.
XX	PR	01-FEB-2000; 2000US-0179531P.
XX	PA	(REGC) UNIV CALIFORNIA.
XX	PI	Eisenberg D, Rotstein SH, Marcotte EM;
XX	DR	WPI; 2001-329193/34.
XX	DR	N-ESDB; AAHS1961.
XX	PT	Identifying nucleotide or polypeptide sequence for use as drug target
XX	PT	involves providing algorithm that analyzes a functional relationship

PT between nucleotide or polypeptide sequences, and comparing the sequences.
 XX Disclosure; Page 157; 207pp; English.

XX This invention relates to a method for identifying a nucleotide or
 CC polypeptide sequence that may be a drug target, or essential for growth
 CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
 CC represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium
 CC tuberculosis proteins which are potential drug targets. The DNA and
 CC protein sequences are used to illustrate the method of the invention. The
 CC method involves providing an unknown nucleotide or polypeptide sequences,
 CC and comparing it to a number of sequences along with at least one
 CC algorithm capable of analysing a functional relationship between
 CC nucleotide and polypeptide sequences. The method is useful for
 CC characterising the function of nucleic acids and polypeptides that may be
 CC useful as a target for a drug or essential for the growth or viability of
 CC an organism

XX Sequence 355 AA;

Query Match 99.4%; Score 1792; DB 4; Length 355;
 Best Local Similarity 99.4%; Pred. No. 4e-125;
 Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSNSRRSLRWSLLSVLAAGVGLATAPACAAPPALSQDRFADFPALPLDPSAMVAQVA 60
 Db 1 MSNSRRSLRWSLLSVLAAGVGLATAPACAAPPALSQDRFADFPALPLDPSAMVAQVG 60
 QY 61 PQVYNINIKLGNNAVAGAGTGVIDPENGVLTTNNHVIAGATDINAFSVGSGQTYGVDVWG 120
 Db 61 PQVYNINIKLGNNAVAGAGTGVIDPENGVLTTNNHVIAGATDINAFSVGSGQTYGVDVWG 120
 QY 121 YDRQDVAVQLRAGAGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 Db 121 YDRQDVAVQLRAGAGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 QY 181 QASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSQGGQGFA 240
 Db 181 QASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSQGGQGFA 240
 QY 241 IPIQAMAIAGQIRSGGSGPTVHIGPTAFLGLGVVDNNGNGARVQRVVGSAAPASLGIST 300
 Db 241 IPIQAMAIAGQIRSGGSGPTVHIGPTAFLGLGVVDNNGNGARVQRVVGSAAPASLGIST 300
 QY 301 GDVITAVDGAIPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
 Db 301 GDVITAVDGAIPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 13

AA04830
 ID AA04830 standard; protein; 379 AA.

XX AC AA04830;

XX 06-JUL-1999 (first entry)

XX Mycobacterium species protein sequence 50F.

XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.

XX Mycobacterium sp.

OS WO9909186-A2.

XX 25-FEB-1999.

XX 14-AUG-1998; 98WO-FR001913.

XX 14-AUG-1997; 97FR-00010404.

XX 11-SEP-1997; 97FR-00011325.

PA (INSP) INST PASTEUR.

XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;

PI Goguet De La Salmnciere Y;

XX WPI; 1999-181045/15.

DR N-PSDB; AAX34252.

XX Mycobacterial DNA vectors containing reporter constructs - for
 CC identifying coding or promoter sequences involved in infection-associated
 CC protein expression.

XX Claim 32; Fig 50F; 309pp; French.

XX Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
 CC from various Mycobacterium species microorganisms. The encoding
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection

XX Sequence 379 AA;

Query Match 99.4%; Score 1792; DB 2; Length 379;
 Best Local Similarity 99.4%; Pred. No. 4.3e-125;
 Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MSNSRRSLRWSLLSVLAAGVGLATAPACAAPPALSQDRFADFPALPLDPSAMVAQVA 60
 Db 25 MSNSRRSLRWSLLSVLAAGVGLATAPACAAPPALSQDRFADFPALPLDPSAMVAQVG 84
 QY 61 PQVYNINIKLGNNAVAGAGTGVIDPENGVLTTNNHVIAGATDINAFSVGSGQTYGVDVWG 120
 Db 85 PQVYNINIKLGNNAVAGAGTGVIDPENGVLTTNNHVIAGATDINAFSVGSGQTYGVDVWG 144
 QY 121 YDRQDVAVQLRAGAGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 180
 Db 145 YDRQDVAVQLRAGAGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVGRVVALGQTV 204
 QY 181 QASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSQGGQGFA 240
 Db 205 QASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSQGGQGFA 264
 QY 241 IPIQAMAIAGQIRSGGSGPTVHIGPTAFLGLGVVDNNGNGARVQRVVGSAAPASLGIST 300
 Db 265 IPIQAMAIAGQIRSGGSGPTVHIGPTAFLGLGVVDNNGNGARVQRVVGSAAPASLGIST 324
 QY 301 GDVITAVDGAIPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
 Db 325 GDVITAVDGAIPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 379

RESULT 14

AAE29702
 ID AAE29702 standard; protein; 330 AA.

XX AC AAE29702;

XX 27-JAN-2003 (first entry)

XX Mycobacterium tuberculosis mature Ra35 antigenic protein.

XX Vaccine; immunity; diagnostic agent; gene therapy; Ra35 antigen.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT Misc-difference 182 /note= "Encoded by GAG"

FT Misc-difference 183 /note= "Encoded by GCG"

XX WO200272792-A2.

```
XX 19-SEP-2002.
XX
XX 13-MAR-2002; 2002WO-US008223.
XX
XX 13-MAR-2001; 2001US-0275837P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Brannon M, Guderian J;
XX
XX WPI; 2002-759844/82.
XX
XX N-PSDB; AAD47077.
XX
XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
XX M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
XX against pathogenic microorganisms e.g. Leishmania and Mycobacterium
XX tuberculosis.
XX
XX Disclosure; Page 79-80; 155pp; English.
XX
XX The invention relates to a recombinant nucleic acid molecule encoding a
XX fusion polypeptide. The recombinant nucleic acid comprises a heterologous
XX polynucleotide sequence encoding an antigen or an antigenic fragment from
XX Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
XX polypeptide or its fragment. The Leishmania polynucleotide is selected
XX from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
XX are used in methods for eliciting immune response in mammals. They are
XX useful as vaccines to elicit protective immunity against pathogenic
XX microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
XX polypeptides are used for enhancing the expression of polynucleotides, as
XX in vivo diagnostic agents and for raising antibodies in a non-human
XX animal. The invention is used in gene therapy. The present sequence is M.
XX tuberculosis mature Ra35 (N-terminus of MTB32A; RA35FL) antigenic protein
XX
XX Sequence 330 AA;
XX
XX Query Match 90.7%; Score 1634; DB 5; Length 330;
XX Best Local Similarity 99.4%; Pred. No. 2.1e-113;
XX Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNAVGAGTGIVDPNGVILT 92
XX |||||
XX 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNAVGAGTGIVDPNGVILT 67
XX |||||
XX
XX 93 NNHVIAGATDINAFSGSGGTGYVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
XX |||||
XX 68 NNHVIAGATDINAFSGSGGTGYVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
XX |||||
XX
XX 153 VVAMNSGGGGGTTPRVPGRVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPDGSGGPV 212
XX |||||
XX 128 VVAMNSGGGGGTTPRVPGRVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPDGSGGPV 187
XX |||||
XX
XX 213 VNLGLQVVGNTAASDNFQSGGCGFAIPIGQAMATAGQIRSGGSGPTVHIGTAFGL 272
XX |||||
XX 188 VNLGLQVVGNTAASDNFQSGGCGFAIPIGQAMATAGQIRSGGSGPTVHIGTAFGL 247
XX |||||
XX
XX 273 GVVDNNGNGARVQVWGSAAPASLIGISTGVDVITAVDQAPINSATAMADALNGHPGDVIS 332
XX |||||
XX 248 GVVDNNGNGARVQVWGSAAPASLIGISTGVDVITAVDQAPINSATAMADALNGHPGDVIS 307
XX |||||
XX
XX 333 VNWQTKSGGTRTGNVTIAEGPPA 355
XX |||||
XX 308 VTWQTKSGGTRTGNVTIAEGPPA 330
XX |||||
XX
XX RESULT 15
XX AAE17566
XX ID AAE17566 standard; protein; 330 AA.
XX
XX AAE17566;
XX
XX 22-APR-2002 (first entry)
XX
XX Mycobacterium sp. MTB32A (Ra35FL) mature protein.
XX
XX Fusion protein; antigen; serological sensitivity; immune response;
XX tuberculosis; infection; vaccine; MTB32A; Ra32FL protein.
XX
XX Mycobacterium sp.
XX
XX Key Location/Qualifiers
XX Region 8..202
XX Misc-difference 182 /note= "Ra35 N-terminal peptide"
XX Misc-difference 183 /note= "Encoded by GAG"
XX Misc-difference 199..330 /note= "Encoded by GCG"
XX Region 199..330 /note= "Ra35 C-terminal peptide, Ra12"
XX
XX WO200198460-A2.
XX
XX 27-DEC-2001.
XX
XX 20-JUN-2001; 2001WO-US019959.
XX
XX 20-JUN-2000; 2000US-00597796.
XX
XX 01-FEB-2001; 2001US-0265737P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Alderson M;
XX
XX WPI; 2002-147798/19.
XX
XX N-PSDB; AAD28336.
XX
XX Composition comprising MTB39 antigen and MTB32A antigen from
XX Mycobacterium species, useful for eliciting immune response in a subject.
XX
XX Claim 70; Fig 6; 136pp; English.
XX
XX The present invention relates to fusion proteins containing at least two
XX Mycobacterium species antigens, nucleotides encoding them and
XX compositions comprising such fusion proteins. The present invention
XX particularly relates to nucleic acids encoding fusion proteins that
XX include two or more individual M. tuberculosis antigens which increase
XX the serological sensitivity of sera from individuals infected with
XX tuberculosis and methods for their use in diagnosis, prevention and
XX treatment of tuberculosis infection. Sequences of the invention are
XX useful for eliciting an immune response in a mammal, e.g., human,
XX immunised with BCG. They are useful in the diagnosis, treatment and
XX prevention of Mycobacterium infection. The fusion proteins and the
XX polynucleotides are useful as diagnostic tools in patients infected with
XX Mycobacterium, in vitro and in vivo assays for detecting humoral
XX antibodies or cell-mediated immunity against M. tuberculosis, for the
XX diagnosis of an infection or monitoring of disease progression, as
XX immunogens to generate or elicit a protective immune response in a
XX patient and for raising anti-M. tuberculosis antibodies in a non-human
XX animal. Sequences of the invention are also used as vaccines. MTB32A
XX fusion proteins of the invention are useful as in vivo diagnostic agents
XX for intradermal skin test. The present sequence is Mycobacterium species
XX MTB32A (Ra32FL) mature protein
XX
XX Sequence 330 AA;
XX
XX Query Match 90.7%; Score 1634; DB 5; Length 330;
XX Best Local Similarity 99.4%; Pred. No. 2.1e-113;
XX Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNAVGAGTGIVDPNGVILT 92
XX |||||
XX 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNAVGAGTGIVDPNGVILT 67
XX |||||
XX
XX 93 NNHVIAGATDINAFSGSGGTGYVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
XX |||||
XX
```

Db 68 NNHVIAGATDINAFVSGSQTYGVVDVGYDRTQDVAVLQLRGAGGLPSAAIGCGVAVGEP 127
QY 153 VVAMGNSGGQGTTPRAVPRVVALGQTVQASDSLGTGABETLNGLIQPDAAIQGDSGGPV 212
Db 128 VVAMGNSGGQGTTPRAVPRVVALGQTVQASDSLGTGABETLNGLIQPDAAIQGDSGGPV 187
QY 213 VNGLQGVVGMNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGSGPTVHIGPTAFGL 272
Db 188 VNGLQGVVGMNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGSGPTVHIGPTAFGL 247
QY 273 GVVDNNGNGARVQVRVVGSAAPASLIGISTGDVITAVDGPINSATAMADALNGHHPGDVIS 332
Db 248 GVVDNNGNGARVQVRVVGSAAPASLIGISTGDVITAVDGPINSATAMADALNGHHPGDVIS 307
QY 333 VNWQTKSGGTRTGNVTLAEGPPA 355
Db 308 VTWQTKSGGTRTGNVTLAEGPPA 330

Search completed: June 22, 2004, 17:15:23
Job time : 44.583 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:07:54 ; Search time 12.6725 Seconds
(without alignments)
1446.225 Million cell updates/sec

Title: US-09-886-349A-2

Perfect score: 1802

Sequence: 1 MSNSRRSLRWLSLLVLA.....QTKSGGTRGNTVLAEGPPA 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PGTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1802	100.0	355	3	US-08-818-112-79
2	1802	100.0	355	4	US-08-818-111-80
3	1802	100.0	355	4	US-09-056-556-79
4	1802	100.0	355	4	US-09-072-596-80
5	1802	100.0	355	4	US-09-072-967-79
6	983	54.6	596	4	US-09-287-849-26
7	983	54.6	729	4	US-09-223-040-2
8	983	54.6	729	4	US-09-287-849-2
9	676	37.5	132	3	US-08-818-112-66
10	676	37.5	132	4	US-08-818-111-67
11	676	37.5	132	4	US-09-056-556-66
12	676	37.5	132	4	US-09-072-596-67
13	676	37.5	132	4	US-09-072-967-66
14	676	37.5	132	4	US-09-636-215-819
15	676	37.5	132	4	US-09-636-215-848
16	676	37.5	132	4	US-09-685-166A-819
17	676	37.5	132	4	US-09-685-166A-848
18	671	37.2	231	4	US-09-287-849-28
19	628	34.9	224	4	US-09-636-215-825
20	628	34.9	224	4	US-09-685-166A-825
21	628	34.9	267	4	US-09-643-597-352
22	628	34.9	267	4	US-09-606-421B-352
23	628	34.9	273	4	US-09-736-457-1864
24	628	34.9	299	4	US-09-643-597-354
25	628	34.9	299	4	US-09-606-421B-354
26	628	34.9	304	4	US-09-636-215-835
27	628	34.9	304	4	US-09-685-166A-835

28	628	34.9	314	4	US-09-736-457-1863
29	628	34.9	400	4	US-09-636-215-852
30	628	34.9	400	4	US-09-685-166A-852
31	628	34.9	487	4	US-09-620-412C-349
32	628	34.9	487	4	US-09-598-419-349
33	628	34.9	518	4	US-09-620-412C-333
34	628	34.9	518	4	US-09-598-419-333
35	628	34.9	525	4	US-09-556-877-196
36	628	34.9	525	4	US-09-620-412C-196
37	628	34.9	525	4	US-09-598-419-196
38	628	34.9	583	4	US-09-620-412C-353
39	628	34.9	583	4	US-09-598-419-353
40	628	34.9	585	4	US-09-620-412C-337
41	628	34.9	585	4	US-09-598-419-337
42	628	34.9	619	4	US-09-620-412C-309
43	628	34.9	619	4	US-09-598-419-309
44	628	34.9	631	4	US-09-620-412C-325
45	628	34.9	631	4	US-09-598-419-325

ALIGNMENTS

RESULT 1
US-08-818-112-79
; Sequence 79, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSER: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-79

Query Match 100.0%; Score 1802; DB 3; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e-151;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSRLRWSLLSVLAAGVGLATAPAAQAPALSDRFPALPLDPSAMVAQVA 60
DB 1 MNSRRSRLRWSLLSVLAAGVGLATAPAAQAPALSDRFPALPLDPSAMVAQVA 60
QY 61 PQVNTNKLGYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVVG 120
DB 61 PQVNTNKLGYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVVG 120
QY 121 YRTOQVAVLQIRGAGGLPSAAIGGVAVGEPVWAMGNSGGGTPRAVPRVVALGQTV 180
DB 121 YRTOQVAVLQIRGAGGLPSAAIGGVAVGEPVWAMGNSGGGTPRAVPRVVALGQTV 180
QY 181 QASDSLTGAETLNGLIQFDDAAIQPGDSGGPVVNGLGQVWGMNTAASDNFOLSGGGGFA 240
DB 181 QASDSLTGAETLNGLIQFDDAAIQPGDSGGPVVNGLGQVWGMNTAASDNFOLSGGGGFA 240
QY 241 IPGQAMAIAGIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAAPASLGIST 300
DB 241 IPGQAMAIAGIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAAPASLGIST 300
QY 301 GDVITAVDGAPINSATAMADALNGHHPGDVSVNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGAPINSATAMADALNGHHPGDVSVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 2

US-08-818-111-80
; Sequence 80, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-80

Query Match 100.0%; Score 1802; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e-151;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSRLRWSLLSVLAAGVGLATAPAAQAPALSDRFPALPLDPSAMVAQVA 60
DB 1 MNSRRSRLRWSLLSVLAAGVGLATAPAAQAPALSDRFPALPLDPSAMVAQVA 60
QY 61 PQVNTNKLGYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVVG 120
DB 61 PQVNTNKLGYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVVG 120
QY 121 YRTOQVAVLQIRGAGGLPSAAIGGVAVGEPVWAMGNSGGGTPRAVPRVVALGQTV 180
DB 121 YRTOQVAVLQIRGAGGLPSAAIGGVAVGEPVWAMGNSGGGTPRAVPRVVALGQTV 180
QY 181 QASDSLTGAETLNGLIQFDDAAIQPGDSGGPVVNGLGQVWGMNTAASDNFOLSGGGGFA 240
DB 181 QASDSLTGAETLNGLIQFDDAAIQPGDSGGPVVNGLGQVWGMNTAASDNFOLSGGGGFA 240
QY 241 IPGQAMAIAGIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAAPASLGIST 300
DB 241 IPGQAMAIAGIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAAPASLGIST 300
QY 301 GDVITAVDGAPINSATAMADALNGHHPGDVSVNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGAPINSATAMADALNGHHPGDVSVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 3

US-09-056-556-79
; Sequence 79, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-79

Query Match 100.0%; Score 1802; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e-151;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSRLRWSLLSVLAAGVGLATAPAAQAPALSDRFPALPLDPSAMVAQVA 60

Db 1 MNSRRSLRWSLLSVLAAGVGLATAPAAAPALSDQRFADFPALPLDPSAMVAQVA 60
QY 61 PQVNTINTKLYNNVAGAGTGIIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDVVG 120
Db 61 PQVNTINTKLYNNVAGAGTGIIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDVVG 120
QY 121 YDRTQDVAVLOLRGAGGLPSAAIGGGVAVGEPVAVMNSGGGGTTPRAVGRVVALGQTV 180
Db 121 YDRTQDVAVLOLRGAGGLPSAAIGGGVAVGEPVAVMNSGGGGTTPRAVGRVVALGQTV 180
QY 181 QASDSLTCGAETLNGLIQFDAAIQPGDGGPVVNGLGQVVGWNTAASDNFQLSQGGQGFA 240
Db 181 QASDSLTCGAETLNGLIQFDAAIQPGDGGPVVNGLGQVVGWNTAASDNFQLSQGGQGFA 240
QY 241 IPGQMAIAGQIRSGGGSPVTHIGTAFGLGVVDNNGNGARVQVVGSAFAASLGIST 300
Db 241 IPGQMAIAGQIRSGGGSPVTHIGTAFGLGVVDNNGNGARVQVVGSAFAASLGIST 300
QY 301 GDVITAVDGPINSATAMADALNGHHPGDVSVNWQTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDGPINSATAMADALNGHHPGDVSVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 4
US-09-072-596-80
; Sequence 80, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-596-80

Query Match 100.0%; Score 1802; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e-151;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAGVGLATAPAAAPALSDQRFADFPALPLDPSAMVAQVA 60
Db 1 MNSRRSLRWSLLSVLAAGVGLATAPAAAPALSDQRFADFPALPLDPSAMVAQVA 60
QY 61 PQVNTINTKLYNNVAGAGTGIIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDVVG 120
Db 61 PQVNTINTKLYNNVAGAGTGIIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDVVG 120
QY 121 YDRTQDVAVLOLRGAGGLPSAAIGGGVAVGEPVAVMNSGGGGTTPRAVGRVVALGQTV 180
Db 121 YDRTQDVAVLOLRGAGGLPSAAIGGGVAVGEPVAVMNSGGGGTTPRAVGRVVALGQTV 180
QY 181 QASDSLTCGAETLNGLIQFDAAIQPGDGGPVVNGLGQVVGWNTAASDNFQLSQGGQGFA 240
Db 181 QASDSLTCGAETLNGLIQFDAAIQPGDGGPVVNGLGQVVGWNTAASDNFQLSQGGQGFA 240
QY 241 IPGQMAIAGQIRSGGGSPVTHIGTAFGLGVVDNNGNGARVQVVGSAFAASLGIST 300
Db 241 IPGQMAIAGQIRSGGGSPVTHIGTAFGLGVVDNNGNGARVQVVGSAFAASLGIST 300
QY 301 GDVITAVDGPINSATAMADALNGHHPGDVSVNWQTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDGPINSATAMADALNGHHPGDVSVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 5
US-09-072-967-79
; Sequence 79, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-967-79

```
Query Match      100.0%; Score 1802; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e-151; Indels 0; Gaps 0;
Matches 355; Conservative 0; Mismatches 0;

QY 1 MSNRRSLRWSLLSLVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
Db 1 MSNRRSLRWSLLSLVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60

QY 61 PQVNNINTKLGNNAVAGAGTIVDPNGVLTNNHVIAGATDINAFSVSGQTYGVDDVVG 120
Db 61 PQVNNINTKLGNNAVAGAGTIVDPNGVLTNNHVIAGATDINAFSVSGQTYGVDDVVG 120

QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGPEVPMNGSGGGGTTPRAVPRVVALGQTV 180
Db 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGPEVPMNGSGGGGTTPRAVPRVVALGQTV 180

QY 181 QASDSLTGABETLNGLIQFDAAIQFGSDSGGPFVNVNGLGVVGMNTAASDNFQLSQGGGPPA 240
Db 181 QASDSLTGABETLNGLIQFDAAIQFGSDSGGPFVNVNGLGVVGMNTAASDNFQLSQGGGPPA 240

QY 241 IPIQMAIAIQIRSGGSPVTHIGPTAFGLGLGVDDNNGNGARVQVVGSAAPASLGIST 300
Db 241 IPIQMAIAIQIRSGGSPVTHIGPTAFGLGLGVDDNNGNGARVQVVGSAAPASLGIST 300

QY 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 6
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1997-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match      54.6%; Score 983; DB 4; Length 596;
Best Local Similarity 97.5%; Pred. No. 1.1e-78;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 APAQAAPPALSDQRFADFPALPLDPSAMVAQVPOVNVNITKLGNNAVAGAGTGIVIDPN 87
Db 397 AAGDIAPPALSDQRFADFPALPLDPSAMVAQVPOVNVNITKLGNNAVAGAGTGIVIDPN 456

RESULT 7
US-09-223-040-2
; Sequence 2, Application US/09223040
; Patent No. 6544522
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-223-040-2

Query Match      54.6%; Score 983; DB 4; Length 729;
Best Local Similarity 97.5%; Pred. No. 1.5e-78;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 APAQAAPPALSDQRFADFPALPLDPSAMVAQVPOVNVNITKLGNNAVAGAGTGIVIDPN 87
Db 530 AAGDIAPPALSDQRFADFPALPLDPSAMVAQVPOVNVNITKLGNNAVAGAGTGIVIDPN 589

QY 88 GVLTTNNHVIAGATDINAFSVSGQTYGVDDVVDRTQDVAVLQLRGAGGLPSAAIGGV 147
Db 590 GVLTTNNHVIAGATDINAFSVSGQTYGVDDVVDRTQDVAVLQLRGAGGLPSAAIGGV 649

QY 148 AVGEPVAMNGSGGGGTTPRAVPRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPGD 207
Db 650 AVGEPVAMNGSGGGGTTPRAVPRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPGD 709

QY 208 SGGPVVNGLGQVVGMMNTAAS 227
Db 710 SGGPVVNGLGQVVGMMNTAAS 729

RESULT 8
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
```

```
/ CURRENT APPLICATION NUMBER: US/09/287,849
/ CURRENT FILING DATE: 1999-04-07
/ PRIOR APPLICATION NUMBER: US 06/818,112
/ PRIOR FILING DATE: 1997-03-13
/ PRIOR APPLICATION NUMBER: US 08/942,578
/ PRIOR FILING DATE: 1997-10-01
/ PRIOR APPLICATION NUMBER: US 09/025,197
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 09/056,556
/ PRIOR FILING DATE: 1998-04-07
/ PRIOR APPLICATION NUMBER: US 09/223,040
/ PRIOR FILING DATE: 1998-12-30
/ NUMBER OF SEQ ID NOS: 46
/ SOFTWARE: Patent In Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 729
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-2

Query Match          54.6%; Score 983; DB 4; Length 729;
Best Local Similarity 97.5%; Pred. No. 1.5e-78;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 APAQAAPPALSDQRFADFPALPLDPSAMVAQAPQVNNINTKLGYNNAYGAGTGIVIDPN 87
Db 530 AAGDIAPPALSDQRFADFPALPLDPSAMVAQAPQVNNINTKLGYNNAYGAGTGIVIDPN 589
QY 88 GVVLTNNHVIAGATDINAFSVGSQGYGVYDVVGYDRTQDVAVLQLRGAGLPSAATGGGV 147
Db 590 GVVLTNNHVIAGATDINAFSVGSQGYGVYDVVGYDRTQDVAVLQLRGAGLPSAATGGGV 649
QY 148 AVGEPVVMGNSGGQGTTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAATQPGD 207
Db 650 AVGEPVVMGNSGGQGTTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAATQPGD 709
QY 208 SGGPVVNGLGQVVGMMTAAS 227
Db 710 SGGPVVNGLGQVVGMMTAAS 729

RESULT 9
US-08-818-112-66
; Sequence 66, Application US/08818112
; Patent No. 6290569
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997

Query Match          37.5%; Score 676; DB 3; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASDNFOLSQGGQFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGAR 283
Db 1 TAASDNFOLSQGGQFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGAR 60
QY 284 VQVVVGSAPAASLGISTGDIVITAVDGPINSATAMADALNGHHPGCDVTSVNWQTKSGGTR 343
Db 61 VQVVVGSAPAASLGISTGDIVITAVDGPINSATAMADALNGHHPGCDVTSVNWQTKSGGTR 120
QY 344 TGNVTLAEGPPA 355
Db 121 TGNVTLAEGPPA 132

RESULT 10
US-08-818-111-67
; Sequence 67, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
```

```

;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-67

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 283
Db 1 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60

QY 284 VQRVVGSAAPASLGISTGDTVAVDGPINSATAMADALNGHHHPGDVISVNWQTKSGGTR 343
Db 61 VQRVVGSAAPASLGISTGDTVAVDGPINSATAMADALNGHHHPGDVISVNWQTKSGGTR 120

QY 344 TGNVTLAEGPPA 355
Db 121 TGNVTLAEGPPA 132

RESULT 11
US-09-556-556-66
; Sequence 66, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; CITY: Seattle
; STATE: Washington
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 622-6031
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-66

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 283
Db 1 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60

QY 284 VQRVVGSAAPASLGISTGDTVAVDGPINSATAMADALNGHHHPGDVISVNWQTKSGGTR 343
Db 61 VQRVVGSAAPASLGISTGDTVAVDGPINSATAMADALNGHHHPGDVISVNWQTKSGGTR 120

QY 344 TGNVTLAEGPPA 355
Db 121 TGNVTLAEGPPA 132

TREATM

;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-67

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 283
Db 1 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60

QY 284 VQRVVGSAAPASLGISTGDTVAVDGPINSATAMADALNGHHHPGDVISVNWQTKSGGTR 343
Db 61 VQRVVGSAAPASLGISTGDTVAVDGPINSATAMADALNGHHHPGDVISVNWQTKSGGTR 120

QY 344 TGNVTLAEGPPA 355
Db 121 TGNVTLAEGPPA 132

RESULT 12
US-09-072-596-67
; Sequence 67, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Meto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-67

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 283
Db 1 TAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60

QY 284 VQRVVGSAAPASLGISTGDTVAVDGPINSATAMADALNGHHHPGDVISVNWQTKSGGTR 343
Db 61 VQRVVGSAAPASLGISTGDTVAVDGPINSATAMADALNGHHHPGDVISVNWQTKSGGTR 120

QY 344 TGNVTLAEGPPA 355
Db 121 TGNVTLAEGPPA 132

RESULT 13
US-09-072-967-66
```

; Sequence 66, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-967-66

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 224 TAASDNFOLSGGCGFAIPGQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGAR 283
Db 1 TAASDNFOLSGGCGFAIPGQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGAR 60

Qy 284 VQVVGSAPAAASLGISTGDTVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTR 343
Db 61 VQVVGSAPAAASLGISTGDTVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTR 120

Qy 344 TGNVTLAEGPPA 355
Db 121 TGNVTLAEGPPA 132

RESULT 14
US-09-636-215-819
; Sequence 819, Application US/09636215
; Patent No. 6620922
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui

; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stoik, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.42717C17
; CURRENT APPLICATION NUMBER: US/09/636,215
; CURRENT FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 852
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 819
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapien
; US-09-636-215-819

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 224 TAASDNFOLSGGCGFAIPGQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGAR 283
Db 1 TAASDNFOLSGGCGFAIPGQMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGAR 60

Qy 284 VQVVGSAPAAASLGISTGDTVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTR 343
Db 61 VQVVGSAPAAASLGISTGDTVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTR 120

Qy 344 TGNVTLAEGPPA 355
Db 121 TGNVTLAEGPPA 132

RESULT 15
US-09-636-215-848
; Sequence 848, Application US/09636215
; Patent No. 6620922
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stoik, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.42717C17
; CURRENT APPLICATION NUMBER: US/09/636,215
; CURRENT FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 852
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 848
; LENGTH: 132

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-636-215-848

Query Match      37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred.No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 224 TAASDNFQLSQGGGFAIPICQAWAIAGQIRSGGSPVHIGPTAFGLGVVDNNGNGAR 283
    |||||
Db 1 TAASDNFQLSQGGGFAIPICQAWAIAGQIRSGGSPVHIGPTAFGLGVVDNNGNGAR 60
    |||||

Qy 284 VQRVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTR 343
    |||||
Db 61 VQRVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTR 120
    |||||

Qy 344 TGNVTLAEGPPA 355
    |||||
Db 121 TGNVTLAEGPPA 132
    |||||
```

Search completed: June 22, 2004, 17:27:28
Job time : 13.6725 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 32.8799 Seconds
(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-2

Perfect score: 1802
Sequence: 1 MSNRRSLRWLSVLA.....QPKSGGTGNTVLAEGPPA 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 28213646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1802	100.0	355	12	US-09-886-349A-2
2	1802	100.0	355	14	US-10-193-002-80
3	1802	100.0	355	14	US-10-084-843-79
4	1802	100.0	355	14	US-10-098-732A-2
5	1792	99.4	355	9	US-09-712-363-161
6	1634	90.7	330	12	US-09-886-349A-4
7	1634	90.7	330	14	US-10-098-732A-4
8	1634	90.7	330	15	US-10-369-983-19
9	1631	90.5	330	12	US-09-886-349A-6
10	1631	90.5	330	14	US-10-098-732A-6
11	1631	90.5	330	15	US-10-369-983-20
12	1631	90.5	723	15	US-10-369-983-2
13	1631	90.5	1010	15	US-10-369-983-4
14	999.5	55.5	1016	15	US-10-369-983-18
15	992	55.0	1154	15	US-10-369-983-16

16	990.5	55.0	825	15	US-10-369-983-14	Sequence 14, Appl
17	989.5	54.9	813	15	US-10-369-983-15	Sequence 15, Appl
18	988	54.8	1022	15	US-10-369-983-17	Sequence 17, Appl
19	985	54.7	195	12	US-09-886-349A-8	Sequence 8, Appl
20	985	54.7	195	14	US-10-098-732A-8	Sequence 8, Appl
21	983.5	54.6	875	15	US-10-369-983-13	Sequence 13, Appl
22	983	54.6	596	9	US-09-287-849-26	Sequence 26, Appl
23	983	54.6	596	12	US-09-886-349A-20	Sequence 20, Appl
24	983	54.6	596	14	US-10-359-460-26	Sequence 26, Appl
25	983	54.6	596	14	US-10-098-732A-20	Sequence 20, Appl
26	983	54.6	729	9	US-09-287-849-2	Sequence 2, Appl
27	983	54.6	729	12	US-09-886-349A-16	Sequence 16, Appl
28	983	54.6	729	14	US-10-359-460-2	Sequence 2, Appl
29	983	54.6	729	14	US-10-098-732A-16	Sequence 16, Appl
30	983	54.6	729	15	US-10-369-983-21	Sequence 21, Appl
31	983	54.6	729	15	US-10-359-459-2	Sequence 2, Appl
32	983	54.6	930	14	US-10-098-732A-65	Sequence 65, Appl
33	983	54.6	930	15	US-10-369-983-12	Sequence 12, Appl
34	980	54.4	729	12	US-09-886-349A-18	Sequence 18, Appl
35	980	54.4	729	14	US-10-098-732A-18	Sequence 18, Appl
36	980	54.4	729	15	US-10-369-983-22	Sequence 22, Appl
37	676	37.5	132	9	US-09-759-143-819	Sequence 819, App
38	676	37.5	132	9	US-09-759-143-848	Sequence 848, App
39	676	37.5	132	9	US-09-780-669-819	Sequence 819, App
40	676	37.5	132	9	US-09-780-669-848	Sequence 848, App
41	676	37.5	132	9	US-09-822-827-819	Sequence 819, App
42	676	37.5	132	9	US-09-822-827-848	Sequence 848, App
43	676	37.5	132	9	US-09-895-793-819	Sequence 819, App
44	676	37.5	132	9	US-09-895-793-848	Sequence 848, App
45	676	37.5	132	9	US-09-895-814-819	Sequence 819, App

ALIGNMENTS

RESULT 1

US-09-886-349A-2
; Sequence 2, Application US/09886349A
; Publication No. US2004008523A1

; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir

; APPLICANT: Reed, Steven

; APPLICANT: Alderson, Mark

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis

; FILE REFERENCE: 014058-00907005

; CURRENT APPLICATION NUMBER: US/09/886,349A

; PRIOR FILING DATE: 2001-06-20

; PRIOR APPLICATION NUMBER: US 09/597,796

; PRIOR FILING DATE: 2000-06-20

; PRIOR APPLICATION NUMBER: US 60/265,737

; PRIOR FILING DATE: 2001-02-01

; NUMBER OF SEQ ID NOS: 50

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 355

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: MTB32A (Ra35FL)

US-09-886-349A-2

Query Match 100.0%; Score 1802; DB 12; Length 355;

Best Local Similarity 100.0%; Pred. No. 1.6e-134;

Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNRRSLRWLSVLA...VGLGLATAPAAAPALSQDRFADFPALPLDPSAMVAQVA 60

Db 1 MSNRRSLRWLSVLA...VGLGLATAPAAAPALSQDRFADFPALPLDPSAMVAQVA 60

QY 61 PQVNVNITKLYNNVAGAGTGVIDP...VGLGLATAPAAAPALSQDRFADFPALPLDPSAMVAQVA 120

Db 61 PQVNVNITKLYNNVAGAGTGVIDP...VGLGLATAPAAAPALSQDRFADFPALPLDPSAMVAQVA 120

QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPGRVVVALGQTV 180
 Db 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPGRVVVALGQTV 180
 QY 181 QASDSLTCABETLNLQFDAAIQPDGSGGPVNVGLGVVGMNTAASDNFOLSGQGQGA 240
 Db 181 QASDSLTCABETLNLQFDAAIQPDGSGGPVNVGLGVVGMNTAASDNFOLSGQGQGA 240
 QY 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVRVVGSAAPAASLGIST 300
 Db 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVRVVGSAAPAASLGIST 300
 QY 301 GDVITAVDGPINSATAMADALNGHHGPDVSVNWQTKSGGTRTGNVTLAEGPPA 355
 Db 301 GDVITAVDGPINSATAMADALNGHHGPDVSVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 2

US-10-193-002-80
 ; Sequence 80, Application US/10193002
 ; Publication No. US20030135026A1
 ; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 ; Skeiky, Yasir A.W.
 ; Dillon, Davin C.
 ; Campos-Neto, Antonia
 ; Houghton, Raymond
 ; Vedvick, Thomas S.
 ; Twardzik, Daniel R.
 ; Lodes, Michael J.
 ; Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 TUBERCULOSIS

NUMBER OF SEQUENCES: 350

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/193.002
 FILING DATE: 10-Jul-2002
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072.596
 FILING DATE: 05-MAY-1998
 ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.417C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 80:

SEQUENCE CHARACTERISTICS:

LENGTH: 355 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 80:

US-10-193-002-80

Query Match 100.0%; Score 1802; DB 14; Length 355;
 Best Local Similarity 100.0%; Pred. No. 1.6e-134;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNRRRLSRWSLLSVLAAVGLGLATAPAAQAPALSQDRFADFPALPLDPSAMVAQVA 60
 Db 1 MSNRRRLSRWSLLSVLAAVGLGLATAPAAQAPALSQDRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNINTKLYNNAVGAGTGIVIDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVDVVG 120
 Db 61 PQVNINTKLYNNAVGAGTGIVIDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVDVVG 120
 QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPGRVVVALGQTV 180
 Db 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPGRVVVALGQTV 180
 QY 181 QASDSLTCABETLNLQFDAAIQPDGSGGPVNVGLGVVGMNTAASDNFOLSGQGQGA 240
 Db 181 QASDSLTCABETLNLQFDAAIQPDGSGGPVNVGLGVVGMNTAASDNFOLSGQGQGA 240
 QY 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVRVVGSAAPAASLGIST 300
 Db 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVRVVGSAAPAASLGIST 300
 QY 301 GDVITAVDGPINSATAMADALNGHHGPDVSVNWQTKSGGTRTGNVTLAEGPPA 355
 Db 301 GDVITAVDGPINSATAMADALNGHHGPDVSVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 3

US-10-084-843-79
 ; Sequence 79, Application US/10084843
 ; Publication No. US20030143243A1
 ; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 ; Skeiky, Yasir A.W.
 ; Dillon, Davin C.
 ; Campos-Neto, Antonio
 ; Houghton, Raymond
 ; Vedvick, Thomas S.
 ; Twardzik, Daniel R.
 ; Lodes, Michael J.
 ; Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/084,843
 FILING DATE: 25-Feb-2002
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072,967
 FILING DATE: 05-MAY-1998

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.411C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 79:

SEQUENCE CHARACTERISTICS:

LENGTH: 355 amino acids

TYPE: amino acid

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-084-843-79

Query Match      100.0%; Score 1802; DB 14; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.6e-134;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAAPPALSQDRFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAAPPALSQDRFPALPLDPSAMVAQVA 60
QY 61 PQVNTNKLGYNNVAGAGTGVIDPNGVVLTNHVIAGATDINAFSVSGGQTYGVDVVG 120
DB 61 PQVNTNKLGYNNVAGAGTGVIDPNGVVLTNHVIAGATDINAFSVSGGQTYGVDVVG 120
QY 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVVMNGSGGGGTTPRAVPRVVALGQTV 180
DB 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVVMNGSGGGGTTPRAVPRVVALGQTV 180
QY 181 QASDSLTCABEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAASDNFOLSQGGQGFA 240
DB 181 QASDSLTCABEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAASDNFOLSQGGQGFA 240
QY 241 IPTQAMAIAGQIRSGGSPTHVIGPTAFGLGVVDNNGNGARVQRVVGSAAPASLGIST 300
DB 241 IPTQAMAIAGQIRSGGSPTHVIGPTAFGLGVVDNNGNGARVQRVVGSAAPASLGIST 300
QY 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAGGPPA 355
DB 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAGGPPA 355
```

```
RESULT 4
US-10-098-732A-2
; Sequence 2, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-0120100S
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,637
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; TYPE: PRT
; LENGTH: 355
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB32A (Ra35FL)
US-10-098-732A-2
```

```
Query Match      100.0%; Score 1802; DB 14; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.6e-134;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAAPPALSQDRFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAAPPALSQDRFPALPLDPSAMVAQVA 60
QY 61 PQVNTNKLGYNNVAGAGTGVIDPNGVVLTNHVIAGATDINAFSVSGGQTYGVDVVG 120
DB 61 PQVNTNKLGYNNVAGAGTGVIDPNGVVLTNHVIAGATDINAFSVSGGQTYGVDVVG 120
QY 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVVMNGSGGGGTTPRAVPRVVALGQTV 180
```

```
DB 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVVMNGSGGGGTTPRAVPRVVALGQTV 180
QY 181 QASDSLTCABEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAASDNFOLSQGGQGFA 240
DB 181 QASDSLTCABEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAASDNFOLSQGGQGFA 240
QY 241 IPTQAMAIAGQIRSGGSPTHVIGPTAFGLGVVDNNGNGARVQRVVGSAAPASLGIST 300
DB 241 IPTQAMAIAGQIRSGGSPTHVIGPTAFGLGVVDNNGNGARVQRVVGSAAPASLGIST 300
QY 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAGGPPA 355
DB 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAGGPPA 355

RESULT 5
US-09-712-363-161
; Sequence 161, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Edward M.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 161
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-161
```

```
Query Match      99.4%; Score 1792; DB 9; Length 355;
Best Local Similarity 99.4%; Pred. No. 9.6e-134;
Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAAPPALSQDRFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWSLLSVLAAGVGLGATAPAAQAAPPALSQDRFPALPLDPSAMVAQVG 60
QY 61 PQVNTNKLGYNNVAGAGTGVIDPNGVVLTNHVIAGATDINAFSVSGGQTYGVDVVG 120
DB 61 PQVNTNKLGYNNVAGAGTGVIDPNGVVLTNHVIAGATDINAFSVSGGQTYGVDVVG 120
QY 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVVMNGSGGGGTTPRAVPRVVALGQTV 180
DB 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVVMNGSGGGGTTPRAVPRVVALGQTV 180
QY 181 QASDSLTCABEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAASDNFOLSQGGQGFA 240
```

Db 181 QASDSLTGAEETLNGLIQFDAAIQPGSGGPPVNVNGLQVVMNTAASDNFQLSQGGGPPA 240
Qy 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGARVQVWGSAPASLGIST 300
Db 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGARVQVWGSAPASLGIST 300
Qy 301 GDTVAVDGPINSATAMADALNGHHPGDIIVSVNWTQKSGGTRTGNVTLAEGPPA 355
Db 301 GDTVAVDGPINSATAMADALNGHHPGDIIVSVNWTQKSGGTRTGNVTLAEGPPA 355

RESULT 6
US-09-886-349A-4
; Sequence 4, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB32A (Ra35 mature)
US-09-886-349A-4

Query Match 90.7%; Score 1634; DB 12; Length 330;
Best Local Similarity 99.4%; Pred. No. 2.9e-121;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 33 APPALSQDRFADFALPLDPSAMVAQVAPQVNVNINIKLGYNNVAVGAGTGIVDPNGVLT 92
Db 8 APPALSQDRFADFALPLDPSAMVAQVAPQVNVNINIKLGYNNVAVGAGTGIVDPNGVLT 67

Qy 93 NNHVIAGATDINAFSVGSGQTYGVVDVGYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEP 152
Db 68 NNHVIAGATDINAFSVGSGQTYGVVDVGYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEP 127

Qy 153 VVMGNSGGGGTTPRVPGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPV 212
Db 128 VVMGNSGGGGTTPRVPGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPV 187

Qy 213 VNLGQVVGMNTAASDNFQLSQGGGPAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGL 272
Db 188 VNLGQVVGMNTAASDNFQLSQGGGPAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGL 247

Qy 273 GVVDNNGGARVQVWGSAPASLGISTGDTVAVDGPINSATAMADALNGHHPGDIIVSVN 332
Db 248 GVVDNNGGARVQVWGSAPASLGISTGDTVAVDGPINSATAMADALNGHHPGDIIVSVN 307

Qy 333 VNWQKSGGTRTGNVTLAEGPPA 355
Db 308 VTWQKSGGTRTGNVTLAEGPPA 330

RESULT 7
US-10-098-732A-4
; Sequence 4, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir

; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Ra35 mature
US-10-098-732A-4

Query Match 90.7%; Score 1634; DB 14; Length 330;
Best Local Similarity 99.4%; Pred. No. 2.9e-121;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 33 APPALSQDRFADFALPLDPSAMVAQVAPQVNVNINIKLGYNNVAVGAGTGIVDPNGVLT 92
Db 8 APPALSQDRFADFALPLDPSAMVAQVAPQVNVNINIKLGYNNVAVGAGTGIVDPNGVLT 67

Qy 93 NNHVIAGATDINAFSVGSGQTYGVVDVGYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEP 152
Db 68 NNHVIAGATDINAFSVGSGQTYGVVDVGYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGEP 127

Qy 153 VVMGNSGGGGTTPRVPGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPV 212
Db 128 VVMGNSGGGGTTPRVPGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPV 187

Qy 213 VNLGQVVGMNTAASDNFQLSQGGGPAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGL 272
Db 188 VNLGQVVGMNTAASDNFQLSQGGGPAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGL 247

Qy 273 GVVDNNGGARVQVWGSAPASLGISTGDTVAVDGPINSATAMADALNGHHPGDIIVSVN 332
Db 248 GVVDNNGGARVQVWGSAPASLGISTGDTVAVDGPINSATAMADALNGHHPGDIIVSVN 307

Qy 333 VNWQKSGGTRTGNVTLAEGPPA 355
Db 308 VTWQKSGGTRTGNVTLAEGPPA 330

RESULT 8
US-10-369-983-19
; Sequence 19, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: wild-type
; OTHER INFORMATION: mature MTB32A (Ra35)

US-10-369-983-19

Query Match 90.7%; Score 1634; DB 15; Length 330;
Best Local Similarity 99.4%; Pred. No. 2.9e-121;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 33 APPALSQDRFADFPALPLDPSAMVAQVPOVNVNINIKLGYNNVAGAGTGVIDPNGVLT 92
DB 8 APPALSQDRFADFPALPLDPSAMVAQVPOVNVNINIKLGYNNVAGAGTGVIDPNGVLT 67

QY 93 NNHVIAGATDINAFSGSGQTYGVYVGYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEP 152
DB 68 NNHVIAGATDINAFSGSGQTYGVYVGYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEP 127

QY 153 VVAMNSGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
DB 128 VVAMNSGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187

QY 213 VNGLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGL 272
DB 188 VNGLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGL 247

QY 273 GVVDNNGGARVQVRVVGSAAPASLSIGSTGCVITAVDGPINSATAMADALNGHHPGDVIS 332
DB 248 GVVDNNGGARVQVRVVGSAAPASLSIGSTGCVITAVDGPINSATAMADALNGHHPGDVIS 307

QY 333 VNWQTKSGGTRTGNVTLAEGPPA 355
DB 308 VNWQTKSGGTRTGNVTLAEGPPA 330

RESULT 9

US-09-886-349A-6
; Sequence 6, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; PRIOR FILING DATE: 2000-06-20
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Ra35FLMutSA
US-09-886-349A-6

Query Match 90.5%; Score 1631; DB 12; Length 330;
Best Local Similarity 99.1%; Pred. No. 4.9e-121;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 33 APPALSQDRFADFPALPLDPSAMVAQVPOVNVNINIKLGYNNVAGAGTGVIDPNGVLT 92
DB 8 APPALSQDRFADFPALPLDPSAMVAQVPOVNVNINIKLGYNNVAGAGTGVIDPNGVLT 67

QY 93 NNHVIAGATDINAFSGSGQTYGVYVGYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEP 152
DB 68 NNHVIAGATDINAFSGSGQTYGVYVGYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEP 127

QY 153 VVAMNSGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
DB 128 VVAMNSGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187

QY 213 VNGLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGL 272
DB 188 VNGLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGL 247

QY 273 GVVDNNGGARVQVRVVGSAAPASLSIGSTGCVITAVDGPINSATAMADALNGHHPGDVIS 332
DB 248 GVVDNNGGARVQVRVVGSAAPASLSIGSTGCVITAVDGPINSATAMADALNGHHPGDVIS 307

QY 333 VNWQTKSGGTRTGNVTLAEGPPA 355
DB 308 VNWQTKSGGTRTGNVTLAEGPPA 330

RESULT 10

US-10-098-732A-6
; Sequence 6, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Gudarian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Ra35FLMutSA
US-10-098-732A-6

Query Match 90.5%; Score 1631; DB 14; Length 330;
Best Local Similarity 99.1%; Pred. No. 4.9e-121;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 33 APPALSQDRFADFPALPLDPSAMVAQVPOVNVNINIKLGYNNVAGAGTGVIDPNGVLT 92
DB 8 APPALSQDRFADFPALPLDPSAMVAQVPOVNVNINIKLGYNNVAGAGTGVIDPNGVLT 67

QY 93 NNHVIAGATDINAFSGSGQTYGVYVGYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEP 152
DB 68 NNHVIAGATDINAFSGSGQTYGVYVGYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEP 127

QY 153 VVAMNSGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
DB 128 VVAMNSGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187

QY 213 VNGLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGL 272
DB 188 VNGLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGL 247

QY 273 GVVDNNGGARVQVRVVGSAAPASLSIGSTGCVITAVDGPINSATAMADALNGHHPGDVIS 332
DB 248 GVVDNNGGARVQVRVVGSAAPASLSIGSTGCVITAVDGPINSATAMADALNGHHPGDVIS 307

QY 333 VNWQTKSGGTRTGNVTLAEGPPA 355
DB 308 VNWQTKSGGTRTGNVTLAEGPPA 330

RESULT 11

US-10-369-983-20
; Sequence 20, Application US/10369983
; Publication No. US20030235593A1

; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Guderian, Jeff
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009081US
 ; CURRENT APPLICATION NUMBER: US/10/369,983
 ; CURRENT FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: US 60/357,351
 ; PRIOR FILING DATE: 2002-02-15
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 20
 ; LENGTH: 330
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:mutated
 ; OTHER INFORMATION: MTB32MutSA (Ra35 mutSA)
 US-10-369-983-20

Query Match 90.5%; Score 1631; DB 15; Length 330;
 Best Local Similarity 99.1%; Pred. No. 4.9e-121;
 Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 33 APPALSQDRFADFPALPLDPSAMVAQVAVPQVNNINTKLGYNNAVAGAGTGIVDPNGVWLT 92
 DB 8 APPALSQDRFADFPALPLDPSAMVAQVAVPQVNNINTKLGYNNAVAGAGTGIVDPNGVWLT 67
 QY 93 NNHVIAGATDINAFSVSGGQTYGVVDVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
 DB 68 NNHVIAGATDINAFSVSGGQTYGVVDVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
 QY 153 VVAMGNSGGGQTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
 DB 128 VVAMGNSGGGQTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDAGGPV 187
 QY 213 VNLGQVVGMMNTAASDNFQLSQGGQGFPAIPGQAMAIAGQIRSGGSPVTHIGPTAFLGL 272
 DB 188 VNLGQVVGMMNTAASDNFQLSQGGQGFPAIPGQAMAIAGQIRSGGSPVTHIGPTAFLGL 247
 QY 273 GVDNNGNGARVQVVGSAAPASISLGISTGDIVITAVDGAPINSATAMADALNGHHPGDVIS 332
 DB 248 GVDNNGNGARVQVVGSAAPASISLGISTGDIVITAVDGAPINSATAMADALNGHHPGDVIS 307
 QY 333 VNMQTKSGGTRTGNVTLAEGPPA 355
 DB 308 VTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 12
 US-10-369-983-2
 ; Sequence 2, Application US/10369983
 ; Publication No. US20030235593A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Guderian, Jeff
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009081US
 ; CURRENT APPLICATION NUMBER: US/10/369,983
 ; CURRENT FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: US 60/357,351
 ; PRIOR FILING DATE: 2002-02-15
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 723
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:mutated
 ; OTHER INFORMATION: MTB32-MTB39F fusion protein (MTB32MutSA)
 US-10-369-983-2

Query Match 90.5%; Score 1631; DB 15; Length 723;
 Best Local Similarity 99.1%; Pred. No. 1.3e-120;
 Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 33 APPALSQDRFADFPALPLDPSAMVAQVAVPQVNNINTKLGYNNAVAGAGTGIVDPNGVWLT 92
 DB 8 APPALSQDRFADFPALPLDPSAMVAQVAVPQVNNINTKLGYNNAVAGAGTGIVDPNGVWLT 67
 QY 93 NNHVIAGATDINAFSVSGGQTYGVVDVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
 DB 68 NNHVIAGATDINAFSVSGGQTYGVVDVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
 QY 153 VVAMGNSGGGQTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
 DB 128 VVAMGNSGGGQTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDAGGPV 187
 QY 213 VNLGQVVGMMNTAASDNFQLSQGGQGFPAIPGQAMAIAGQIRSGGSPVTHIGPTAFLGL 272
 DB 188 VNLGQVVGMMNTAASDNFQLSQGGQGFPAIPGQAMAIAGQIRSGGSPVTHIGPTAFLGL 247
 QY 273 GVDNNGNGARVQVVGSAAPASISLGISTGDIVITAVDGAPINSATAMADALNGHHPGDVIS 332
 DB 248 GVDNNGNGARVQVVGSAAPASISLGISTGDIVITAVDGAPINSATAMADALNGHHPGDVIS 307
 QY 333 VNMQTKSGGTRTGNVTLAEGPPA 355
 DB 308 VTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 13
 US-10-369-983-4
 ; Sequence 4, Application US/10369983
 ; Publication No. US20030235593A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Guderian, Jeff
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009081US
 ; CURRENT APPLICATION NUMBER: US/10/369,983
 ; CURRENT FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: US 60/357,351
 ; PRIOR FILING DATE: 2002-02-15
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 4
 ; LENGTH: 1010
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:MTB-102F fusion
 ; OTHER INFORMATION: protein
 US-10-369-983-4

Query Match 90.5%; Score 1631; DB 15; Length 1010;
 Best Local Similarity 99.1%; Pred. No. 2e-120;
 Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 33 APPALSQDRFADFPALPLDPSAMVAQVAVPQVNNINTKLGYNNAVAGAGTGIVDPNGVWLT 92
 DB 8 APPALSQDRFADFPALPLDPSAMVAQVAVPQVNNINTKLGYNNAVAGAGTGIVDPNGVWLT 67
 QY 93 NNHVIAGATDINAFSVSGGQTYGVVDVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
 DB 68 NNHVIAGATDINAFSVSGGQTYGVVDVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
 QY 153 VVAMGNSGGGQTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212

Db 128 VVAMGNSGGGTPRAVPGRVVVALGQTVQASDLSLTGAETLNLGTFDAATQPGDAGGPV 187
QY 213 VNLGGOVGMNTAASNFOLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFILQL 272
Db 188 VNLGGOVGMNTAASNFOLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFILGL 247
QY 273 GVVDDNNGARVQVRVVGSAASLSLTGSDVITAVDGPINSATAMADALNGHHFPGDVIS 332
Db 248 GVVDDNNGARVQVRVVGSAASLSLTGSDVITAVDGPINSATAMADALNGHHFPGDVIS 307
QY 333 VNMOTKSGGTRGNVTLAEGPPA 355
Db 308 VTMOTKSGGTRGNVTLAEGPPA 330

RESULT 14
US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85b)
US-10-369-983-18

Query Match 55.5%; Score 999.5; DB 15; Length 1016;
Best Local Similarity 84.1%; Pred. No. 2e-70;
Matches 207; Conservative 6; Mismatches 18; Indels 15; Gaps 4;
QY 28 APAQAAPPALSDQRFADFPALPLDPSAMVAQVAPQVNVNINIKLGYNNNAVAGTGIVIDPN 87
Db 530 AAGDIAPPALSDQRFADFPALPLDPSAMVAQVAPQVNVNINIKLGYNNNAVAGTGIVIDPN 589
QY 88 GVLTNNHVIAGATDINAFSVGSGQTYGVDVVDRTQDVAVLQLRGAGGLPSAAIGGGV 147
Db 590 GVLTNNHVIAGATDINAFSVGSGQTYGVDVVDRTQDVAVLQLRGAGGLPSAAIGGGV 649
QY 148 AVGEPVVMGNSGGGTPRAVPGRVVVALGQTVQASDLSLTGAETLNLGTFDAATQPGD 207
Db 650 AVGEPVVMGNSGGGTPRAVPGRVVVALGQTVQASDLSLTGAETLNLGTFDAATQPGD 709
QY 208 SGGPVVNLGQVGMNTAASNFOLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPT 267
Db 710 SGGPVVNLGQVGMNTAASGTF--SRPG----LPVEYLQVPSMGRDIXVQFQSGENN 763
QY 259 SPTVHI 264
Db 764 SPAVYL 769

RESULT 15
US-10-369-983-16
; Sequence 16, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff

; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB114F (MTB72F-mTCC#2)
US-10-369-983-16

Query Match 55.0%; Score 992; DB 15; Length 1154;
Best Local Similarity 63.8%; Pred. No. 9.2e-70;
Matches 219; Conservative 17; Mismatches 59; Indels 48; Gaps 6;
QY 28 APAQAAPPALSDQRFADFPALPLDPSAMVAQVAPQVNVNINIKLGYNNNAVAGTGIVIDPN 87
Db 530 AAGDIAPPALSDQRFADFPALPLDPSAMVAQVAPQVNVNINIKLGYNNNAVAGTGIVIDPN 589
QY 88 GVLTNNHVIAGATDINAFSVGSGQTYGVDVVDRTQDVAVLQLRGAGGLPSAAIGGGV 147
Db 590 GVLTNNHVIAGATDINAFSVGSGQTYGVDVVDRTQDVAVLQLRGAGGLPSAAIGGGV 649
QY 148 AVGEPVVMGNSGGGTPRAVPGRVVVALGQTVQASDLSLTGAETLNLGTFDAATQPGD 207
Db 650 AVGEPVVMGNSGGGTPRAVPGRVVVALGQTVQASDLSLTGAETLNLGTFDAATQPGD 709
QY 208 SGGPVVNLGQVGMNTAASNFOLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPT 267
Db 710 SGGPVVNLGQVGMNTAASGTFMDF-----GLLPP-----EVSNSMYSGGPESMLAAA 760
QY 268 AFLGLGVVDNNGARVQVRVVGSAASLSLTGSDVITAVDGPIT--NSATAMADALNGH 325
Db 761 AWDGV-----AELTSAVSYGSVSVSTLIVEPMMGPAAAAAATAATPY 803
QY 326 HPGDVISVNW-----QTKSGGTRGNVTLAEGPPA 355
Db 804 -----VGLAATAALAKETATQAAAAAFAFGTAFAMTVPPS 839

Search completed: June 22, 2004, 18:07:51
Job time : 34.8799 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:59:04 ; Search time 9.93246 Seconds
(without alignments)
3438.018 Million cell updates/sec

Title: US-09-886-349A-2
Perfect score: 1802
Sequence: 1 MNSRRSLRWLSVLAA.....QTKSGGTRGNVTLAEGPPA 355

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*

1: Piri: *
2: Piri: *
3: Piri: *
4: Piri: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1792	99.4	355	2 F70983	probable serine pr
2	1306	72.5	361	2 S47170	hypothetical prote
3	1272.5	70.6	354	2 A87242	probable secreted
4	468.5	28.0	464	2 C70821	probable serine pr
5	460	25.5	382	2 H86930	probable secreted
6	460	25.5	452	2 T45448	probable serine pr
7	382	21.2	394	2 S74643	proteinase rhoA (E
8	372	20.6	407	2 AG2150	serine proteinase
9	355.5	19.7	452	2 S77538	serine proteinase
10	354	19.6	433	2 H97199	htrA-like serine p
11	345	19.1	362	2 T35257	probable secreted
12	343	19.0	457	2 AG0433	proteinase (EC 3.4
13	341.5	19.0	514	2 A82591	periplasmic protei
14	335.5	18.6	429	2 AD1894	serine proteinase
15	331	18.4	416	2 AB2057	serine proteinase
16	331	18.4	441	2 B75357	probable periplasm
17	331	18.4	455	2 C91142	serine endoprotein
18	331	18.4	455	2 F85987	serine endoprotein
19	329.5	18.3	474	2 F83550	serine proteinase
20	329	18.3	398	2 B71284	probable periplasm
21	329	18.3	455	2 AB0909	serine protease (E
22	328	18.2	355	1 JG6052	trypsin-like prote
23	328	18.2	355	2 D91142	trypsin-like prote
24	327	18.1	352	2 B82307	trypsin-like prote
25	327	18.1	355	2 G85987	trypsin-like prote
26	326	18.1	455	2 JG6051	heat shock protein
27	322	17.9	475	1 S15337	protease DO precur
28	319	17.7	475	2 AC0528	proteinase DO (EC
29	317.5	17.6	474	2 S45229	

RESULT 1

F70983

probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002

C:Accession: F70983

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whithead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID:98295987; PMID:9634230

A: Accession: F70983

A: Status: Preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-355 <COL>

A: Cross-references: GB:Z96071; GB:AL123456; NID:g3242254; PIDN:CAR09453.1; PID:g2181957

A: Experimental source: strain H37RV

C: Geneticks:

A: Gene: pepA

C: Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp

Query Match 99.4%; Score 1792; DB 2; Length 355;

Best Local Similarity 99.4%; Pred. No. 7.5e-98;

Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNSRRSLRWLSVLAAVGLGLATAPAAQAPALQSDRFPADFPALPLDPSAMVAQVA 60

Db 1 MNSRRSLRWLSVLAAVGLGLATAPAAQAPALQSDRFPADFPALPLDPSAMVAQVG 60

QY 61 PQVNVNTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGGTGYVDVVG 120

Db 61 PQVNVNTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGGTGYVDVVG 120

QY 121 YRTQDVAVLQIRGAGLPSAAGGVAVGESVWAMNSGGGGTTPRVPGRVWALQTV 180

Db 121 YRTQDVAVLQIRGAGLPSAAGGVAVGESVWAMNSGGGGTTPRVPGRVWALQTV 180

QY 181 QASDSLTAETETLNGLIQFDAAIQPDSGGPVYVGLGVVGMNTAASDNFQLSQGGQGFA 240

Db 181 QASDSLTAETETLNGLIQFDAAIQPDSGGPVYVGLGVVGMNTAASDNFQLSQGGQGFA 240

QY 241 IFPGQAMATAGIRSGGSPVTHIGTAPFLGLGVVDNNGNGARVORVVGSAFAASLIGIST 300

Db 241 IFPGQAMATAGIRSGGSPVTHIGTAPFLGLGVVDNNGNGARVORVVGSAFAASLIGIST 300

QY 301 GDVITAVDCAPINSATAMADALNGHHPQDVISVNWOTKSGGTRTGNVTLAEGPPA 355

Db 301 GDVITAVDCAPINSATAMADALNGHHPQDVISVNWOTKSGGTRTGNVTLAEGPPA 355

30 317.5 17.6 474 2 E85500 proteinase DO (EC
31 317.5 17.6 474 2 E90649 proteinase DO (EC
32 316.5 17.6 348 2 H96956 serine protease Do
33 315.5 17.5 474 1 I40059 htrA-like protein
34 315.5 17.5 474 2 A13349 proteinase DO (EC
35 313.5 17.4 530 2 F87590 serine proteinase
36 312.5 17.3 356 2 SC0909 trypsin-like prote
37 311.5 17.3 416 2 S75445 proteinase rhoB (E
38 311 17.3 401 2 AD2451 serine proteinase
39 310 17.2 408 2 H86891 exported serine pr
40 308.5 17.1 481 2 AH0410 global stress requ
41 306 17.0 472 2 C87408 serine proteinase
42 306 17.0 513 2 AD3418 proteinase DO (EC
43 305 16.9 328 1 D69109 serine proteinase
44 303 16.8 513 2 I40060 serine proteinase
45 303 16.8 499 2 B81914 probable periplasm

ALIGNMENTS

RESULT 2
S47170
hypothetical protein 34K - Mycobacterium paratuberculosis
C:Species: Mycobacterium paratuberculosis
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Dec-2002
C:Accession: S47170
R:Camaron, R.M.; Stevenson, K.; Inglis, N.F.; Klausen, J.; Sharp, J.M.
submitted to the EMBL Data Library, June 1993
A:Description: Isolation and characterisation of a 34KDa protein of Mycobacterium paratuberculosis
A:Reference number: S47170
A:Accession: S47170
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-361 <CAM>
A:Cross-references: EMBL:Z23092; NID:G505550; PIDN:CAA80638.1; PID:G505551
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Query Match 72.5%; Score 1306; DB 2; Length 361;
Best Local Similarity 70.4%; Pred. No. 1.9e-69;
Matches 255; Conservative 41; Mismatches 58; Indels 8; Gaps 3;
QY 1 MSNS-RRRSRLWSLLSVLAAGVGLG-----ATAPAAQAPALSDQRFADFPALPLDPS 53
Db 1 MSKSHHRSVWWSLVGLVGLGSGVGLAPASAPSGUALDRFADRLPLDPS 60
QY 54 AMVAQVAPQVYNINIKLYNNNAVAGTGVIVDPNGVLTNNHVIAGATDINAFSVSGQT 113
Db 61 AMVQVQVQVYNIDTKFYNNNAVAGTGVIVDPNGVLTNNHVIAGTATISAFDVGNGQT 120
QY 114 YGVDMVGVDRTOQVAVLQLEGAGLPSAATGGVAVGEPVAMGNSGGQGTFRAPVGRV 173
Db 121 YAVDMVGVDRTOQVAVLQLEGAGLPSAATGGVAVGEPVAMGNSGGQGTFRAPVGRV 180
QY 174 VALGQTVQASDLSLTGABETLNGLIQFDAAIQPGDSGGPVVNGLVQVVGMMNTAASDNFQLS 233
Db 181 VALNQSVSATDTLTLGAENGLLQADAPIKPGDSGGPVVNGLVQVVGMMNTAATDYSKMS 240
QY 234 QGGGFAIPICQAMAIAGQIRSGGSGTIVHIGPTAFGLGVNDNGNGARVQRVWGSA 293
Db 241 -GGGFAIPICRAMAVANQIRSGGSGTIVHIGPTAFGLGVNDNGNGARVQRVWGSA 299
QY 294 ASLGISTGDTVITAVDGAIPINSATAMADALNGHHPGDVIVSNWOTKSGTGTGNTVLAEGP 353
Db 300 AAGIAPGDTVITGVTPPINGATSMTEVLVPHFGDTIAVHFSVDGGERTANITLAEGP 359
QY 354 PA 355
Db 360 PA 361

RESULT 3
A87242
Probable secreted serine proteinase [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002
C:Accession: A87242
R:Cole, S.T.; Eiglmeyer, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holroyd, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, R.; Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squares, S.; Taylor, K.; Whitehead, S.; Barrell, B.G.
Nature 409, 1007-1011, 2001
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86509; MUID:21128732; PMID:11234002
A:Accession: A87242
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-354 <STO>
A:Cross-references: GB:AL450380; NID:g13093863; PIDN:CAC32191.1; GSPDB:GN00147
C:Genetics:
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Query Match 70.6%; Score 1272.5; DB 2; Length 354;
Best Local Similarity 70.1%; Pred. No. 1.7e-67;
Matches 251; Conservative 42; Mismatches 58; Indels 7; Gaps 3;
QY 1 MSNSRRRSRLWSLLSVLAAGVGLGATAPAAQ---APPALSDQRFADFPALPLDPSAMVA 57
Db 1 MSRQPHRSRLWSLLSVLAAGVGLGATAPAAQ---APPALSDQRFADFPALPLDPSAMVA-- 58
QY 58 QVAPQVYNINIKLYNNNAVAGTGVIVDPNGVLTNNHVIAGATDINAFSVSGQGTGV 117
Db 59 -VAPQVYNINIKLYNNNAVAGTGVIVDPNGVLTNNHVIAGATDINAFSVSGQGTGV 117
QY 118 VVGYDRTOQVAVLQLEGAGLPSAATGGVAVGEPVAMGNSGGQGTFRAPVGRVWALG 177
Db 118 VVGYDRTOQVAVLQLEGAGLPSAATGGVAVGEPVAMGNSGGQGTFRAPVGRVWALN 177
QY 178 QTVQASDLSLTGABETLNGLIQFDAAIQPGDSGGPVVNGLVQVVGMMNTAASDNFQLS 237
Db 178 QTVQASEPLTGAQETLSGLIQVDAPIKPGDSGGPVVNSRGQVVGMMNTAATDNYKN-LGGQ 236
QY 238 GFAPICQAMAIAGQIRSGGSGTIVHIGPTAFGLGVNDNGNGARVQRVWGSA 297
Db 237 GFAPICQAMAIAGQIRSGGSGTIVHIGPTAFGLGVNDNGNGARVQRVWGSA 296
QY 298 ISTGDTVITAVDGAIPINSATAMADALNGHHPGDVIVSNWOTKSGTGTGNTVLAEGPPA 355
Db 297 ISVGDITISVDGVPISSEATAMTNLVPHPHGETVAVNYSAGGGDLTANVLAEGPPA 354

RESULT 4
C70821
probable serine proteinase Rv0983 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002
C:Accession: C70821
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: C70821
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-464 <COL>
A:Cross-references: GB:AL021999; GB:AL123456; NID:G3261536; PIDN:CAA17582.1; PID:e125391
A:Experimental source: strain H37Rv
C:Genetics:
C:Gene: Rv0983
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Query Match 26.0%; Score 468.5; DB 2; Length 464;
Best Local Similarity 36.0%; Pred. No. 2.1e-20;
Matches 127; Conservative 59; Mismatches 136; Indels 31; Gaps 10;
QY 19 AAVGLGLATAPAAQAPALSDQRFADFPALPLDPSA---MVAQVAPQVYNINIKLYNN 75
Db 120 AASLVGNFAPGSGGPVAAASAPSPIPANMPGSGVEQVAAKVVPSVVMLETDLGRQSE 179
QY 76 VVGYDRTOQVAVLQLEGAGLPSAATGGVAVGEPVAMGNSGGQGTFRAPVGRVWALG 127
Db 180 -EGSGIILSAEGLILNNHVIAGATDINAFSVSGQGTGVIVDPNGVLTNNHVIAGATDINAFSVSGQGTGV 237
QY 128 AVLQRLGAGLPSAATGG---VAVGEPVAMGNSGGQGTFRAPVGRVWALGQTVQASD 185
Db 238 AVVRVQVSGGLTPISLGSSDLRVGQVLAIGSLGLEGT--VTTGIVSALNREVSTIGE 295
QY 186 LTGABETLNGLIQFDAAIQPGDSGGPVVNGLVQVVGMMNTA---ASDNFQLSQQ--CQGF 239
Db 296 -AGNQNTVLDAIQDAAINFGNSGALVNNNAQLVGVNSAIATLGADSAQAQSSIGLGF 354
QY 240 AIPICQAMAIAGQIRSGGSGTIVHIGPTAFGLGV--DNNNGARVQRVWGSA 298

C:Genetics:

A:Note: MUCB373.28

C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; trypt

C:Keywords: hydrolase; serine proteinase

F:182,224,305/Active site: His, Asp, Ser #status predicted

Query Match 25.5%; Score 460; DB 2; Length 452;

Best Local Similarity 36.4%; Pred. No. 6.3e-20;

Matches 130; Conservative 52; Mismatches 121; Indels 54; Gaps 12;

QY 22 GLGLATAPAAAPPALSDQRFADFPALPLDPSAMVAQA-----PQVNTNTKLGNNAVG 77

DB 120 GAGPVTGPAAASVPAANM-----PSSGEQAVAKVVPVVMLETLGRQSE-- 164

QY 78 AGTGIVIDPQNGVLTNNHVIAGA-----TDINAFSVSGSGQTYGVDDVGVYDR 123

DB 165 EGSVLIUSADGLIITNNHVAVAAPKGGPGGLSPKTVTF--DGTASTVTVGADP 221

QY 124 TDQVAVTLQRGAGGLPSAAICGG--VAVGEPVVAWNGSGQGCTPRAVPRVVALGQTIVQ 181

DB 222 TSDIADVVRQVSGISGLTPTMGSSADLRVGPVAVGSPGLAGT--VTSGIVSALNRPSV 279

QY 182 ASDSLTGAETLNGLIQFDALIQCDSGGPFVNGIGOVVGMNTAA-----SDNFOLSQG 235

DB 280 TTGE-SGNQNTVLDALQIDAAINPNSGALVNMGGQLVNGVNSAIATLGADSGDAQSGSI 338

QY 236 GQGFAPITQGAAMAIAGQIRSGGSPSTVHIGETAFGLGLGVVDNNGN-GARVQRVVGSAAPAA 294

DB 339 GLGFAPVDAQRIADELISGT--KATH----ASLGVQVATDKGTGCAKVMVDVAGAAA 392

QY 295 SLGISTGDTVAVDQAPINSATAMADALNGHPGDIVSNWQTKSGGTGNTVLAELAE 351

DB 393 NAAVPKGVLTQVDRLLISSDALVAARSKAPGDKVSLTYQDQSGSSRTVQVTLGK 449

RESULT 7

S74643

proteinase hhoA (EC 3.4.-.-) - Synchocystis sp. (strain PCC 6803)

N:Alternate names: protein sll1679

C:Species: Synchocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Dec-2002

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; o. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis s.

A:Reference number: S74322; MUID:97061201; PM-ID:8905231

A:Accession: S74643

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-394 <KAN>

A:Cross-references: EMBL:D90900; GB:AB001339; NID:g1651768; PIDN:BAAL6795.1; PID:g165186

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Gene: hhoA

C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; trypt

C:Keywords: hydrolase; proteinase

Query Match 21.2%; Score 382; DB 2; Length 394;

Best Local Similarity 31.6%; Pred. No. 2e-15;

Matches 116; Conservative 50; Mismatches 117; Indels 84; Gaps 14;

QY 13 WLLSVLAAGVGLATAP-----AQARPPA-----LSQDRFADFPALPLDPSAWVA 57

DB 13 YLLAFVAGTAFGANLPHAVAAADLPAPVITAQASVPLTSESFV-----AAAVS 63

QY 58 QVAPQWNNTK-----LQYNNAV-----GAGTGIVIDPQNGV 90

DB 64 RSGPAVVDIETVTRTDPIILDPPFFQBFGRSPVPRERRIAGQSGGIFDINSGLI 123

QY 91 LTNNHVIAGADINAFSVSGQTYGVDDVGVYDRTDQVAVLQLRGAG-GLPSAAIG--CGV 147

```

Db 124 LTNAHVVDGASKV-VTLRDGRTDQVGRGTDEVTDLAVKIEPQSGALPVAPLGTSSNL 182
QY 148 AVGEPVAVMNGSGGGTTPRAVGRVVALG--QTVQASDSLTGAETLNGLIQFDAAIQP 205
Db 183 QVGDWAIAGVNPVGLDNT-----VTLGIISTLGRSAAQAGIPDKRVEFIQTDAAINP 234
QY 206 GDSGGPVVNGLGQVGMATASDNFQLSQGGFAIPGQAWATAGQIRSGGGSPVTHIG 265
Db 235 GNSGQPLNAGVEVIGINTA-----IRADATGIGFAIPIDQAKAIONTLAAGTVPHPHYIG 290
QY 266 PTAFLGLGV-----VDNNGN-----GARVQVVGSAAPASLGIISTGDVITAVDGP 311
Db 291 -VQMMNITVDQAQQNRNPNPFFIPEVDGILVMKVLPGTPAERAGIRRGDVIIVAVDGP 349
QY 312 INSATAM 318
Db 350 ISDGARL 356

```

RESULT 8

```

AG2150
serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG2150
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
C:Accession: AG2150
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-407 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074457.1; PID:g17131851; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Gene: alt2758
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

```

```

Query Match 20.6%; Score 372; DB 2; Length 407;
Best Local Similarity 30.3%; Pred. No. 7.8e-15;
Matches 125; Conservative 59; Mismatches 141; Indels 88; Gaps 15;

QY 7 RSLR--MSWLLSLVLAAGL---GLATAPAAAP-----PALSDQRFADFPALPLD 51
Db 10 RSRQLGTHVLAFIGVVLTVSSLRVLPSCAEPAPNPSTTGSAPELVAQSQSTAAVIGN 69
QY 52 PS---AMVAQVAVQVNNKLGNNV-----GAGTG 81
Db 70 SSFVTAANVVGSAVVRIDTERTITRRVDPLEDPFRFRFGFGQQLPPEQMRGLGSG 129
QY 82 IVIDPNGVLTNNHVIAGATINAFSGQTVGVVDVDRDQVAVLQLRGAGLPSA 141
Db 130 FIDKSLIILTNHVVDKARVTV-RLKGRSFDGKVGIDVTDVAVKINAGNSLPVA 188
QY 142 AIG--GGVAVGEPVAVMNGSGGGTTPRAVGRVVALG--QTVQASDSLTGAETLNGLI 197
Db 189 PLGSSNNVQVGDWAIAGVNPGLGPDNT-----VTLGIVSTLKRSSAQVGITDKELDFI 240
QY 198 QFDAAIQPGSGGVVNGLGQVGMATASDNFQLSQGGFAIPGQAWATAGQIRSGG 257
Db 241 QTDAAINFGNSGGLNDKGEVIGINTA-----IRADAMGIGFAIPIDKAKATQIERDG 296
QY 258 GSPVTHIGTAFGLGV-----VDNNG-----NGARVQVVGSAAPASLGI 298
Db 297 --KVAH-----PYLGQVQMATLTPELAQQNNIDPNSAFAPVENVGLVIRVVPNSPAANAGI 350
QY 299 STGDVITAVDGAIPNSATAMADALNGHHPGDVTSVNWQTKSGGTRTGNVTLAE 351
Db 351 RRGDVLQVDGQAITAEQLQNVVNSRLGQALQVRLQ-RGNQTQQLSVRTAE 402

```

RESULT 9

```

S77538
serine proteinase (EC 3.4.21.-) htrA - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein sir1204
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Dec-2002
C:Accession: S77538
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
S.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S77538
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-452 <KAN>
A:Cross-references: EMBL:D90905; GB:AB001339; NID:g1652360; PIDN:BA017385.1; PID:g165246;
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try
C:Keywords: hydrolase; serine proteinase

```

```

Query Match 19.7%; Score 355.5; DB 2; Length 452;
Best Local Similarity 29.2%; Pred. No. 8.1e-14;
Matches 116; Conservative 65; Mismatches 129; Indels 87; Gaps 16;

QY 16 SVLAAGVGLGATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQV----APQVNNINTKLG 71
Db 84 SVISPLVTNQSIAPANESLATNLQSRLS-----PREPSNFVVDVVESTGPAVVRAINAQKT 138
QY 72 YNNV-----GAGTGIVIDPNGVLTNNHVIAGATDIN 104
Db 139 VKSQVPQAFNPFFLQRFQSGQMPMPNERVQRTGSGFVSNQDKIFTNAHVVDGADVT 198
QY 105 AFSVSGSGQTVGVVDVDRDQVAVLQLRGAGLPSAAIGGG--VAVGEPVAVMNGSGGQ 162
Db 199 V-TLKGRSFGRVNGSDPSTDVAVVKIE-AGDLPFTVALGSDHLQVGEWAIAIGNPLGL 256
QY 163 GGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVGM 222
Db 257 DNT--VTTGILSATGR--RSAD--IGVPDKRVEFIQTDAAINPNSGGPLNADGQVIGM 310
QY 223 NTAASDNFQLSQGGGFAIPGQAWATAGQIRSGGSPVTHIGTAFGLGVVDNN--- 278
Db 311 NTAIQNAQ-----GIGFAIPINKAQEIAQQLIATG--KVEH-----ATLGQMTMTTELQ 360
QY 279 -----NGGARVQVVGSAAPASLGIISTGDVITAVDGAIPNSATAMADALNGH 325
Db 361 SQIRQETGMNTPDKGVVIMQVMPNSPAATAKLQSGDVLQSLQGPVENAEQVQSLVKGL 420
QY 326 HPGDVISVNWQTKSGGTRTG---NVTLAEG-----PP 354
Db 421 AVGDEVEL-----GILRNQQQLNLTVTIGALPSAPP 451

```

RESULT 10

```

H97199
htrA-like serine protease (with PDZ domain) [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Dec-2002
C:Accession: H97199
R:Noelling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:121359325
C:Accession: H97199
A:Status: preliminary

```

A:Molecule type: DNA
A:Residues: 1-433 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK80387.1; PID:G15025449; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Query Match 19.6%; Score 354; DB 2; Length 433;
Best Local Similarity 32.0%; Pred. No. 9.4e-14;
Matches 101; Conservative 51; Mismatches 110; Indels 54; Gaps 10;

Qy 50 LPSAMVAQVAPVNNITKL-----GYNNAVAGTGIVIDPNGVLT 92
Db 122 LTVSQIVKXVPAVVGSTKTTVTQNDFFSFGSSNGSGSTQEGMGSGIIFNNDGYILT 181
Qy 93 NHVIAAGATDINAFSGSGTGVVGYDRDQVAVLQRLGAGGLPSAAIGGVA--V 149
Db 182 NHVIVKAGDKI-AVILNNKKEVSAKVNNYDEANDIAVIXTSFTVPGVAELGSSASLV 240
Qy 150 GEPVWAMNSGGGGTFRAPVGRVVALGQTQVQASDLSLTGABETLNGLIQFDDAIQPGDSG 209
Db 241 GDSVVAIGNPLGKEFLGTGTGTVVSAVNEVAVSE---GQKQT---YIQTDAINPGNSG 294
Qy 210 GPVWNGLGQVGNWNTAASDNFQSQGQ---GFAIPIGQAMAIAGQIRSGGSPVTHGP 266
Db 295 GPDVNSFGQVVGINS-----KISENGVEGIGFSIFIDTVKSIQNLK-----P 339
Qy 267 TAPLGLG--VVD-----NNGGARVQVVGSAAPASLGISTGVDITAVDGAIPINSATA 317
Db 340 ILMGTSGEAVDKSTAEQHNIPGVVIEQIDPFSSAKAGMQVGVITKFDGKKTSTSD 399
Qy 318 MADALNGHHPGVISV 333
Db 400 IDSIGKXNSGDTVQV 415

RESULT 11
T35287
A:Cross-references: EMBL:AL096972; PIDN:CABS1255.1; GSPDB:GN00070; SCOEDB:SC5F7.30
A:Experimental source: strain A3(2)
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Query Match 19.1%; Score 345; DB 2; Length 362;
Best Local Similarity 22.6%; Pred. No. 2.6e-13;
Matches 125; Conservative 49; Mismatches 154; Indels 56; Gaps 14;

Qy 1 MNSRRSLRMSWLSVLAAVGLGATA-----PQAAPPALSQDR 41
Db 1 MDTSRTRLRLAPVAVPACV-LLLATGCSAGATDRGSGSAREGDTAQAAPRAASEL 59
Qy 42 FADFPALPDPASMAQVAPQVNNITKLGYNNNAVAGTGIVIDPNGVLTNNHVIAGAT 101
Db 60 EADY-----ERVIKDPLSVVQIQ-----GDSLGSGVVYDDKGVVTVNAHVVG--- 103
Qy 102 DINAFSGSGTGVG---VDVVGYDRDQVAVLQRLGAGGLPSA--AIGGGVAGVPVVA 155
Db 104 DAKSFRVTTARTGALTAKLVSSYPEDQLAVIKLDKVPKGMRAARFADSAKVEGVILA 163
Qy 156 MNSGGGGTFRAPVGRVVALGQTQVQASDLSLTGABETLNGLIQFDDAIQPGDSGGPVNG 215

164 MGSPGLGLSS--VTQGIVSATGRVTTEGSGGGTGATIANMVQTSAAINPNSGALVNL 221
Qy 216 LGQVVGWNTAASDNFQSQG---GQFAIPIGQAMAIAGQIRSGQ---GSPVTHIGPTAF 269
Db 222 DQVIGIPTLTAATDPLGDSNAPGIGFAIPASMTVTTVAGQIVRDKGKTVDSRAALGITA- 280
Qy 270 LGLGVDDNNGN--GARVQVVGSAAPASLGISTGVDITAVDGAIPINSATAVADALNGHHP 327
Db 281 --RTVDDSYRPAAGAAVVEVSDGGAADDAGLRPGDVLVGLGDTDTITTTISSEALASMRP 338
Qy 328 GDVIVSNWQTSKSGGTRTGNVTLAE 351
Db 339 GDRTKVTY-TRDGKHTAEVTLGE 361

RESULT 12
AG0433
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C:Accession: AG0433
R:Packhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
geno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, H.
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AG0433
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-457 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC92795.1; PID:G15981488; GSPDB:GN00175
C:Genetics:
A:Gene: degQ
C:Superfamily: Helicobacter serine proteinase
C:Keywords: hydrolase; serine proteinase

Query Match 19.0%; Score 343; DB 2; Length 457;
Best Local Similarity 31.2%; Pred. No. 4.4e-13;
Matches 122; Conservative 53; Mismatches 126; Indels 90; Gaps 18;

Qy 12 SWLLSVLA-AVGLGIATAP---AQAAPPALSQDRFADFPALPDPASMAQVAPQVNNI- 66
Db 5 SLLLSALAI SVGLGLASVPMVSAALPAVAGQ---TLPSL-----APLKVLPVAVSVH 57
Qy 67 -----NTKLGYNNA---VGAGTGIVID-PNGVVLNNHVIAGAT 101
Db 58 VSGSAQQOQLPEEFKFFFGPNAPSGKESRPFEGLOGSVIINAEKGYILTNHVINNAD 117
Qy 102 DINAFSGSGTGVGVDVVGVDRTQDVAVLQRLGAGGLPSAAI--GGGVAGVGPVAVMGS 159
Db 118 KIRV-QINDGREYDAKILGRDEQTDIALQLTDAKNLTAIKIADSDNLRVGDFAVAGNP 176
Qy 160 GQGGTTPRAVGRVVALGQTQVQASDLSLTGAE-ETLNGLIQFDDAIQPGDSGGPVVNGLQ 218
Db 177 FGLGQT--ATSGIISALGR-----SGLNLEGIENFIQTDASINRNSGALVLDGE 226
Qy 219 VVGWNTAASDNFQSQGQ---GFAIPIGQAMATA-----GOIRSG-----GSPVTH 263
Db 227 LIGINTAI-----LAPGGNIGIGFAIPSNMAQLNSQOLIIEFGEVKRGLGIRGSEMTAD 281
Qy 264 IGPTAFLGLGVDDNNGGARVQVVGSAAPASLGISTGVDITAVDGAIPINSATAMADALN 323
Db 282 IAKAFNI-----DAQRGAFVSEVLPKSAAAKAGIKPGDVLISVDGKKISSFAELRAKV 335
Qy 324 GHHPGDVLSVWQTKSGGTRTGNVTLAEPPP 354
Db 336 TTGPGKTIKIG-----LLREGKP 353

RESULT 13
AB2581

periplasmic proteinase XP2241 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: A82581
 R:anonymous. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; PMID:20365717; PMID:110910347
 A:Note: For a complete list of authors see reference number A59328 below
 A:Accession: A82581
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-514 <STM>
 A:Cross-references: GB:A8004037; GB:A8003849; NID:g9107394; PIDN:AAF85040.1; GSPDB:GN001
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H
 as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.A.
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silva
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XP2241
 C:Superfamily: Helicobacter serine proteinase

Query Match 19.0%; Score 341.5; DB 2; Length 514;
 Best Local Similarity 31.6%; Pred. No. 6.1e-13;
 Matches 123; Conservative 55; Mismatches 128; Indels 83; Gaps 17;
 QY 3 NSRRSLRWMLSV---LAAVGLGLATAQAAP--PALSDQRFADFPALPL-----DP 52
 Db 2 NSRIHT-RCFGLLAITPLAACGQHNSPTTAPSTPIVTP---STTPAQVLVAGLPDF 57
 QY 53 SAMVAQAQPVVINT-----KLGYNNVAV----- 76
 Db 58 TQLVDQVGPVNIETVITRKVKGRGIPLNDIDPEFFRRFFGPDQFQMPNQRGQDDDE 117
 QY 77 -----GAGTGIVDPNGVLTNNHVIATGATDINAFSGSGQTYGVVDVYDRTQDVAVL 130
 Db 118 GGAGRGWGSGFIISKDGYLTNNHVIATGASEV-TKLTDRREFKAKIGSDSYQDVALL 176
 QY 131 QLRGAGGLPSAAGGGVAV--GEPVAMGNSGGGGTTPRVPGRVVALGQVQASDSLGT 188
 Db 177 KI-DAKNLPTVRIGDSSLSKGGVVAIGSPFGLDHSVTA--GIVSALGRS-----TS 226
 QY 189 ABEETLNGLIQFDAAIQDSDGSPVWNLGQVGNMNTAASDNFOLSQ--GQGFAPIGQA 246
 Db 227 DQQRVYVFIQDVFQINQSGGGLNTRGEVGIN---SQIFSASGYMGISFAIPINIA 283
 QY 247 MAIAGQIRSGG---GSFTVHIGTAPL---GLGVVDNNGNGARVORVVGVSAPASLGIS 299
 Db 284 INAAEQIRKTKGVQSRMLGVEIGPIDAKAQLGLPDS--RGALVNIPPHSPAAKAGIE 341
 QY 300 TGDVITAVDGAAPNSATAMADALNGHHPG 328
 Db 342 VGDVIRSWGKVISSFSDLPPLIGMMPGP 370
 RESULT 14
 AD1894
 serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AD1894

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AD1894
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-429 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA000019; PID:g17130047; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr0702
 C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; tryp
 Query Match 18.6%; Score 335.5; DB 2; Length 429;
 Best Local Similarity 29.2%; Pred. No. 1.1e-12;
 Matches 110; Conservative 59; Mismatches 137; Indels 71; Gaps 14;
 QY 18 LAAVGLGLATAQAAPPALSDQRFADFPALPLDPSAMVAQAQVNVNNTKLGVNNAV- 76
 Db 67 LAQKASDLAVSRVDAAPP-LGNNTDPNF-----VTQVQVRGPAVVRIEASRTVTSKLP 119
 QY 77 -----GAGTGIVDPNGVLTNNHVIATGATDINAFSGSG 111
 Db 120 AEFNDPFFRRFFGSQLPQOQERVQRTGSGFLISADGSLTNAHVVDGADTVRI-LKDG 178
 QY 112 QTYGVVDVYDRTQDVAVLQRLGAGLPSAAGGG--VAVGEPVAMGNSGGGGTTPRAV 169
 Db 179 RSFGKVLGTDNLTDVAVVKIQ-ANNLPTLVGNSDQLQPCQWAIATGNPLGLDNT--VT 235
 QY 170 PGRVVALGQVQASDSLGTGABETLNGLIQFDAAIQDSDGSPVWNLGQVGNMNTAASDN 229
 Db 236 TGIISATGRT-----SNQIGAPDKRVEYIQTDAALNPGNSGGPLNRYRGEVGMNTAIQ 291
 QY 230 FQLSGGGGFAIPIGQAQVAIAGQIRSGSGSPTHIGTAFGL-----GVV 275
 Db 292 AQ---GLGFAIPKTAQRISNQLIATGKVQHPVLG-IQVGLTPQIKQINSDPNSGLT 346
 QY 276 DNNGAGAVQVRVGSAPASLIGSTGQDVITAVDGAAPNSATAMADALNGHHPGDIVSN- 334
 Db 347 VDRDKGVLVVRVLPNSPAARAGLRAGDVIQKLNGQAVTDASNVORAVENAQVGLQLLEL 406
 QY 335 WTKSGGTRTG-NVTLLA 350
 Db 407 W-----RNGRNLNLA 416
 RESULT 15
 AB2057
 serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AB2057
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AB2057
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-416 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA000019; PID:g17131098; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all2008
 C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; tryp
 Query Match 18.4%; Score 331; DB 2; Length 416;
 Best Local Similarity 31.5%; Pred. No. 2e-12;
 Matches 92; Conservative 57; Mismatches 115; Indels 28; Gaps 10;

Search completed: June 22, 2004, 17:24:47
Job time : 10.9325 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:46:13 ; Search time 5.9081 Seconds
(without alignments)
3128.737 Million cell updates/sec

Title: US-09-886-349A-2

Perfect score: 1802

Sequence: 1 MSNSRRRSURWSLLSVLAA.....QTKSGGTRTGNVTLAEGPPA 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	350.5	19.5	458	1	YVTA_BACSU
2	328	18.2	355	1	DEGS_ECOLI
3	326	18.1	455	1	DEGO_ECOLI
4	322	17.9	475	1	DEGP_SALTY
5	321	17.8	413	1	HTRA_LACHE
6	317.5	17.6	474	1	DEGP_ECOLI
7	310	17.2	408	1	HTRA_LACLA
8	306	17.0	488	1	DEGP_BRUME
9	305	16.9	513	1	DEGP_BRUSU
10	295.5	16.4	449	1	HTRA_BACSU
11	295	16.4	466	1	HTRA_HAEIN
12	291	16.1	504	1	DEGP_RHIME
13	290	16.1	503	1	DEGP_BARHE
14	289.5	16.1	437	1	DEGI_ARATH
15	288	16.0	497	1	DEGP_CHLTR
16	286.5	15.9	497	1	DEGP_CHLMU
17	286	15.9	478	1	DEGP_BUCAI
18	283	15.7	508	1	DEGP_RICCN
19	282	15.6	488	1	DEGP_CHLPN
20	276.5	15.3	400	1	YVTA_BACSU
21	276	15.3	448	1	DEGS_ARATH
22	272.5	15.1	478	1	DEGP_BUCAP
23	263	14.6	458	1	HRA2_HUMAN
24	258	14.3	513	1	DEGP_RICPR
25	254	14.1	458	1	HRA2_MOUSE
26	244.5	13.6	340	1	HRA3_HAEIN
27	238.5	13.2	453	1	HRA3_HUMAN
28	235	13.0	480	1	HRA1_HUMAN
29	231.5	12.8	476	1	HRA4_HUMAN
30	225.5	12.5	460	1	HRA3_MOUSE
31	225	12.5	480	1	HRA1_MOUSE
32	178.5	9.9	321	1	SPPA_ARATH
33	173	9.6	630	1	Y4BT_RHISN

ALIGNMENTS

RESULT 1

ID	YVTA_BACSU	STANDARD;	PRT;	458 AA.
AC	Q9R9IL; Q35021; Q35039;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Probable serine protease yvta (EC 3.4.21.-).			
GN	YVTA OR YVTB OR BSU33000.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
ON	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
EX	MEDLINE=20158875; PubMed=10692364;			
RA	Noone D., Howell A., Devine K.M.,			
RT	Expression of ykdA, encoding a Bacillus subtilis homologue of HtrA,			
RL	is heat shock inducible and negatively autoregulated."			
RL	J. Bacteriol. 182:1592-1599(2000).			
ON	[2]			
RP	SEQUENCE FROM N.A.			
RC	MEDLINE=98015415; PubMed=9353931;			
EX	Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.,			
RT	"Sequencing of regions downstream of addA (98 degrees) and citG (289			
RL	degrees) in Bacillus subtilis."			
RL	Microbiology 143:3305-3308(1997).			
ON	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
EX	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Boursier L., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,			
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,			
RA	Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,			
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,			
RA	Gauche P., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,			
RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,			
RA	Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,			
RA	Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,			
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,			
RA	Noore D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,			
RA	Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,			
RA	Presecan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,			
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie F.,			
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,			
RA	Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,			
RA	Sorokin A., Tacconis E., Takagi T., Takahashi H., Takemaru K.,			
RA	Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,			
RA	Tosato V., Uchiyama S., Vanderbol M., Vannier F., Vassarotti A.,			

34	162.5	9.0	457	1	PRTC_STRGR
35	158.5	8.8	747	1	ELS_BOVIN
36	153	8.5	801	1	Y747 MYCTU
37	150.5	8.4	515	1	Y140 MYCTU
38	147	8.2	299	1	PRTB_STRGR
39	146.5	8.1	864	1	ELS_RAT
40	145	8.0	914	1	WA22 MYCTU
41	142.5	7.9	957	1	Y278 MYCTU
42	141.5	7.9	710	1	FIB1_ADEG1
43	141	7.8	2021	1	OMPA_RICCN
44	139.5	7.7	778	1	YQ34 MYCTU
45	139.5	7.7	2364	1	PGCA_BOVIN

P52320	streptomyc
P04985	bos taurus
Q53810	mycobacteri
Q50594	mycobacteri
P00777	streptomyc
Q93772	rattus norv
O06794	mycobacteri
P36877	mycobacteri
Q64761	avian adeno
Q52657	tickettisia
P71933	mycobacteri
P13608	bos taurus


```

RC RX MEDLINE=0157:H7 / RIMD 0509952;
RA MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Iehii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP PRELIMINARY SEQUENCE OF 82-355 FROM N.A.
RX MEDLINE=88105815; PubMed=3322223;
RA Vogel R.F., Entian K.-D., Mecke D.;
RT "Cloning and sequence of the mdh structural gene of Escherichia coli
RT coding for malate dehydrogenase.";
RL Arch. Microbiol. 149:36-42(1987).
RN [7]
RP IDENTIFICATION
RA Bazan J.F., Pieterick R.J.;
RT "Structural and catalytic models of trypsin-like viral proteases.";
RL Semin. Virol. 1:311-322(1990).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U15661; AAC43993.1; -
DR EMBL; U32495; AAC44006.1; -
DR EMBL; U18997; AAC458037.1; -
DR EMBL; A5000402; AAC76267.1; -
DR EMBL; A5005551; AAG58363.1; -
DR EMBL; A5002564; BAB37531.1; -
DR EMBL; M24777; -; NOT_ANNOTATED_CDS.
DR PIR; D91142; D91142.
DR MEROPS; S01.275; -.
DR EcoGene; EG11652; hhoB.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Serine protease; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 28
FT CHAIN 29 355
FT DOMAIN 281 326 PDZ.
FT ACT_SITE 96 96 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 126 126 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 201 201 CHARGE RELAY SYSTEM (POTENTIAL).
FT CONFLICT 253 253 R -> A (IN REF. 6).
FT CONFLICT 307 307 V -> E (IN REF. 6).
SQ SEQUENCE 355 AA; 37581 MW; D091B4D65E8FE1CC CRC64;

Query Match
Best Local Similarity 30.7%; Pred.No. 1.1e-11; Length 355;
Matches 110; Conservative 47; Mismatches 135; Indels 66; Gaps 13;

QY 15 LSVLAQVGLG-----LATAPQAAPPALSDRFADFPALPDPSAMVAQVAPQVWIN 67
Dd 3 VKLLRSVAIGLIVGALLVAMPRLSLNPLSPQFSDTETPASVNIAYVRAAPVNV- 61
QY 68 TKLGYNNAVGA-----GTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGV 116

```

```

Db 62 -----YNRGLNTNSHNOLEIRTLGSGVIMDQGVIIITNKHVINDAQI-IVALQDGRVFEA 116
QY 117 DVVGYDRTPQVAVLOLRGAGGLPSAIGGGA--VGEPPVAMGNSGGQGTTPRAVPGRVV 174
Db 117 LLVGSDSLTDLAVLKINATGGLTIPINARVPHIGDVVLAIGNP-----Y 162
QY 175 ALGQTV-QASDSLUGAETLN-----GLIOFDAIOPGDSGGPVVNGLGQVGMNTRASD 228
Db 163 NLGQITQTGIISATG-RIGLNPTGRQNFLOTDSINHNSGGALVNSLGLMGTINTLSFD 221
QY 229 --NPQLSGGGGGAIP-----IGQAMALAGQIRSGGSGPTVHIGTAFGLGVV 275
Db 222 KSDGETPEIGFAIPQLATKMKLIRGVRVIGIGR-----EIAPLHAQGGGI- 276
QY 276 DNNNGGARVQVVGSAPAASLGISTGVITAVDPAPINSATAMADALNHHHPGVISV 333
Db 277 -DQLQGIWVNEVSPDGPAAAGIQVNDLIISVDNKPALSALETMDQVAEIRPGSVIEV 333

RESULT 3
DEQG_ECOLI
ID DEQG_ECOLI STANDARD; PRT; 455 AA.
AC P39099;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protease degQ precursor (EC 3.4.21.-).
GN DEQG OR HHOA OR B3234.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Bass S., Gu G., Goddard A.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
[2]
RN RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-32.
RC STRAIN=K12 / W3110;
RX MEDLINE=96165272; PubMed=8576051;
RA Waller P.R., Sauer R.T.;
RT "Characterization of degQ and degS, Escherichia coli genes encoding
RT homologs of the DegP protease.";
RL J. Bacteriol. 178:1146-1153(1996).
[3]
RN RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blather F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -!- FUNCTION: PROTEASE WITH A SHARED SPECIFICITY WITH DEGP/DEGP.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U15661; AAC43992.1; -
DR EMBL; U32495; AAC44005.1; -
DR EMBL; U18997; AAC458036.1; -
DR EMBL; A5000402; AAC76266.1; -

```

```

DR PIR; JC6051; JC6051.
DR MEROPS; S01.274; -.
DR SWISS-2DPAGE; P39099; COLI.
DR EcoGene; EGI2612; degQ.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SMO0228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
DR Hydrolase; Serine protease; Periplasmic; Repeat; Signal;
KW Complete proteome.
FT SIGNAL 1 27
FT CHAIN 28 455
FT DOMAIN 258 349
FT DOMAIN 355 447
FT ACT_SITE 109 109
FT ACT_SITE 139 139
FT ACT_SITE 214 214
SQ SEQUENCE 455 AA; 47205 MW; 6A090F93AC021C83 CRC64;

Query Match 18.1%; Score 326; DB 1; Length 455;
Best Local Similarity 30.4%; Pred. No. 1.9e-11;
Matches 119; Conservative 61; Mismatches 117; Indels 94; Gaps 19;

QY 14 LLSVLA-AVGLGL-ATAPQAAPPALSSQDRFADFPALPLDPSAMVAQVAVQVNVNT--- 68
DB 7 LLSALALSUGLISASQFQVAVIPGVADQ-AFLPSL-----APMLEKVLPAVSVURVEGT 61
QY 69 -----KLGYNNAV-----GAGTIVIDPN-GVLTNNHVIAGTNDINAFS 107
DB 62 ASQGQKIPKFFKFGDLDQPAQPFGLGSGVILNASKGYVLTNNHVIINQAQKI-SIQ 120
QY 108 VSGSGTVGVVDVGVDRDQVAVLQRLGAGGLPSAAIGGG--VAVGPPVVMNGSGGQGT 165
DB 121 LNDGREFDAKLISDQSDIALLIQINPSKLTQIATADSKLVKRGFPAVAVGNPFLGQT 180
QY 166 PRAVPGRVVAGTQVQASDSLTAQE-ETLNGLIQFDAAIQPSDGGPVPVNGLGQVVGWMT 224
DB 181 --ATSGIVSALGR-----SGLNLEGLNFITQDASINRNGSGGALLNGLGELIGINT 230
QY 225 AASDNFOLSOGGQ---GFAIPICQAWATA-----CQIRSGGSPVTHIGTAPLGLGV 274
DB 231 AI-----LAPGGGSGVIGFAIPSNMARTLAQQLIDFGEIKRG-----LGIK 273
QY 275 VDNNGN-----GARVQRVWGSAPAASLIGISTGVDVITAVDGAIPNSATAMADALN 323
DB 274 TENSADIKAFKALDVQKGFVSVLPQSGSAKAGVKAGDIITSLNCKPLNSFAELRSRIA 333
QY 324 GHHPGDVISVNWQTKSGGTRTG-----NVTLL 349
DB 334 TTEFGT-----KVKLGLLENGKPLEVEVTL 358

RESULT 4
ID -DEGP_SALTY STANDARD; PRT; 475 AA.
AC P26982;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protease do precursor (EC 3.4.21.-).
GN DSGP OR HTRA OR PRD OR STM209.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C5;

```

```

RX MEDLINE=91251770; PubMed=1645840;
RA Johnson K., Charles I., Dougan G., Pickard D., O'Gaora P., Costa G.,
RA Ali T., Miller I., Hormaeche C.;
RT "The role of a stress-response protein in Salmonella typhimurium
RT virulence.";
RL Mol. Microbiol. 5:401-407(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Bante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -!- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. SHARED
CC SPECIFICITY WITH HHOA/DEGO.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- INDUCTION: By heat shock.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X54548; CAA38420.1; -.
CC EMBL; AE008704; AAL19173.1; -.
CC PIR; S15337; S15337.
CC MEROPS; S01-273; -.
CC StyGene; SG10173; degP.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ; 2.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SMO0228; PDZ; 2.
CC PROSITE; PS0106; PDZ; 2.
CC Hydrolase; Serine protease; Heat shock; Periplasmic; Repeat; Signal;
KW Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 475
FT DOMAIN 281 372
FT DOMAIN 378 467
FT ACT_SITE 132 132
FT ACT_SITE 162 162
FT ACT_SITE 237 237
FT DISULFID 84 96
FT SEQUENCE 475 AA; 49315 MW; 86E685BF3CIA289F CRC64;

Query Match 17.9%; Score 322; DB 1; Length 475;
Best Local Similarity 29.1%; Pred. No. 3.2e-11;
Matches 118; Conservative 55; Mismatches 126; Indels 106; Gaps 16;

QY 20 AVGLGLATAP--AQAAPPALSSQDRFADFPALPLDPSAMVAQVAVQVNVNT----- 68
DB 12 ALSGLALSPLSATATSSAMTAQQPSL-----APMLEKVPVSVINVEGTTVWTP 67
QY 69 -----KLGYNNAVGA-----GTGIVID-PNGVV 90
DB 68 RMPRNFQFFQDDSPFCQDGSFQNSPFCCGGNGGNGGQQKFMALGSGVIIDAKGYV 127
QY 91 LTNHVIAGTNDINAFVSGSGTVGVVDVGVDRDQVAVLQRLGAGGLPSAAIGGVA-- 148

```

Db 128 VTNHVVNDASVVKV-QLSDGRKFDKAVGVGKDPKSDIALIQNPKNLTAKLADSDLR 186
QY 149 VGEPPVAMNSGGGGTPEAVGRVVALGQTVQASDSLGTGAB-ETLNGLIQFDALQFGD 207
Db 187 VGDYVVAICNPFELGET--VTSIGVSLGR-----SGLNVENYFIOTDAINRGN 236
QY 208 SGGPPVNGLGQVVGWNTA--ASDNFQLSGGGGQGFAP-----IGQAMAIAGQIRSGGGS 259
Db 237 SGGALVNLNGELIGINTAILAFDGGNI--GIGFALPSNMVKNLTSQMVVEYGOVKRG-- 290
QY 260 PTVHIGTFAFLGLGVVDNNGN-----CARVQRVVGSPAPASLIGSTGDVITA 306
Db 291 -----ELGNTGTELSELAKAMKVDAQRCAFSQVMPNPSAAKAGIKAGDVITS 339
QY 307 VDGAPINSATAMADALNGHHPGDEVISVNWQTKSGGTRTGNVTLAE 351
Db 340 LNKGPISSPAALRAQVGTMPVSKISLG-LLREGKAITVNLLELQ 383

RESULT 5

ID HTRA LACHE STANDARD; PRT; 413 AA.
AC Q9Z4H7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine protease do-like htrA (EC 3.4.21.-).
GN HTRA.
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1587;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=53/7;
RX MEDLINE=99047559; PubMed=9829922;
RA Smeds A., Varmanen P.K., Palva A.M.;
RT Molecular characterization of a stress-inducible gene from
Lactobacillus helveticus.;
RL J. Bacteriol. 180:6143-6153(1998).
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ005672; CAA06668.1; -.
DR MEROPS; S01.273; -.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASERS2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS01056; PDZ; 1.
KW Hydrolase; Serine protease; Transmembrane.
FT TRANSMEM 20 40
FT DOMAIN 141 302 CATALYTIC.
FT DOMAIN 303 401 PDZ.
FT ACT_SITE 150 150 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 180 180 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 258 258 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 413 AA; 42647 MW; B16B677991C88707 CRC64;

Query Match 17.8%; Score 321; DB 1; Length 413;
Best Local Similarity 31.3%; Pred. No. 3.2e-11;
Matches 89; Conservative 50; Mismatches 121; Indels 24; Gaps 8;
QY 67 NTKLGYNNVAGCTGIVI---DPNGVVLNNHVIAGATDINAFSVGSGQTVGVVDVGYDR 123
Db 119 SSRNGKLETSESGVVMKSNKGIVTNNHVISGSDAVQVLI-LANGKTVNAKVGKDS 177
QY 124 TDQVAVLQLRGAGGLPSAAIIGG--VAVGPPVAMNSGGGGTTPRAVPRVVALGQTVQ 181
Db 178 TTDLAVLSIDAKVVTQAFQGDGSKHLEAGQTVIAGVSPGLGSEYASTVQGIISAPARTIS 237
QY 182 ASDSLTGAETLNGLIQFDAAIOPGDSGGPVNGLGOVVGWNT--AASDNFQLSGGGGF 239
Db 238 TS---SGNQOT---VIQTDAAINPGSGGALVNSAGQVIGINSMKLQAQSSDGTSGVEGNAF 291
QY 240 AIPGQAMAIAGQIRSGG--GSPTVHIGTFAFLG-----LGVDNNGNGARVQRVWG 289
Db 292 AIPSNVTVTVNELVKKGKITRQLGVRVIALQIGPEGRSLKIKSNLKNIGVIAFVSR 351
QY 290 SAPAASLGISTGVTAVDGAIPNSATAMADALNGHHPGDEVISV 333
Db 352 NGSANAGIKSGDVITKVDGKVEDVASLHSLIYSHKVGDTNVV 395

RESULT 6

ID DEGP_ECOLI STANDARD; PRT; 474 AA.
AC P09376; P15724;
DT 01-MAR-1989 (Rel. 10, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protease do precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR PTD OR B0161 OR Z0173 OR ECS0165.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89057448; PubMed=3057437;
RA Lipinska B., Sharma S., Georgopoulos C.;
RT "Sequence analysis and regulation of the htrA gene of Escherichia
coli: a sigma 32-independent mechanism of heat-inducible
transcription.";
RL Nucleic Acids Res. 16:10053-10067(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
the 2.4-4.1 min (110,917-193,643 bp) region.";
RL Nucleic Acids Res. 22:1637-1639(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W31655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen B., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kaiman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

```

RN RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Hayek G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimallanta E.T., Potamouisis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RL "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [6]
RN RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
RA "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [7]
RN RP SEQUENCE OF 1-50 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=30323597; PubMed=2165018;
RA Quirk S., Bhattacharya S.K., Bessman M.J.;
RT "Primary structure of the deoxyguanosine triphosphate
RT triphosphohydrolase-encoding gene (dgt) of Escherichia coli.";
RL Gene 89:13-18(1990).
RN [8]
RN RP SEQUENCE OF 1-15 FROM N.A.
RC MEDLINE=90207293; PubMed=2157212;
RA Wurgler S.M., Richardson C.C.;
RT "Structure and regulation of the gene for dGTP triphosphohydrolase
RT from Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2740-2744(1990).
RN [9]
RN RP SEQUENCE OF 27-39, AND CHARACTERIZATION.
RX MEDLINE=90202693; PubMed=2180903;
RA Lipinska B., Zylicz M., Georgopoulos C.;
RT "The HtrA (DegP) protein, essential for Escherichia coli survival at
RT high temperatures, is an endopeptidase.";
RL J. Bacteriol. 172:1791-1797(1990).
RN [10]
RN RP SEQUENCE OF 27-30.
RC STRAIN=K12 / W3110;
RX MEDLINE=98263247; PubMed=9600841;
RA Wilkins M.R., Gasteliger E., Tonella L., Ou K., Tyler M.,
RA Sanchez J.-C., Gooley A.A., Waleh B.J., Bairoch A., Appel R.D.,
RA Williams K.L., Hochstrasser D.F.;
RT "Protein identification with N and C-terminal sequence tags in
RT proteome projects.";
RL J. Mol. Biol. 278:599-608(1998).
RN [11]
RN RP IDENTITY OF HTRA AND PROTEASE DO.
RX MEDLINE=9122240; PubMed=2025286;
RA Seol J.H., Woo S.K., Jung E.M., Yoo S.J., Lee C.S., Kim K.J.,
RA Tanaka K., Ichihara A., Ha D.B., Chung C.H.;
RT "Protease Do is essential for survival of Escherichia coli at high
RT temperatures. Its identity with the htra gene product.";
RL Biochem. Biophys. Res. Commun. 176:730-736(1991).
RN [12]
RN RP DISULFIDE BOND.
RX MEDLINE=22760311; PubMed=12878036;
RA Skorko-Glonek J., Zurawa D., Tanfani F., Scire A., Wawrzynow A.,
RA Narkiewicz J., Bertoli E., Lipinska B.;
RT "The N-terminal region of HtraA heat shock protease from Escherichia
RT coli is essential for stabilization of HtraA primary structure and
RT maintaining of its oligomeric structure.";
RL Biochim. Biophys. Acta 1649:171-182(2003).
CC -I- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.

```

```

CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IT CAN DEGRADES
CC ICA, ADA, CASEIN AND GLOBIN. SHARED SPECIFICITY WITH DEGP.
CC -I- SUBUNIT: Multimeric.
CC -I- SUBCELLULAR LOCATION: Periplasmic.
CC -I- INDUCTION: By heat shock.
CC -I- MISCELLANEOUS: HTRA IS INDISPENSABLE FOR BACTERIAL SURVIVAL AT
CC TEMPERATURES ABOVE 42 DEGREES CELSIUS.
CC -I- SIMILARITY: Belongs to peptidase family S2c.
CC -I- SIMILARITY: Contains 2 PDZ/DHR domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M36536; AAA23394.1; -
CC EMBL; X12457; CAA30997.1; -
CC EMBL; D26562; BAB96738.1; -
CC EMBL; AE000125; AAC73272.1; -
CC EMBL; U70214; AAB08591.1; -
CC EMBL; AE005192; AAG54465.1; -
CC EMBL; AF002550; BAB33588.1; -
CC EMBL; M29555; AAA23717.1; -
CC EMBL; M31772; AAA23680.1; -
CC PIR; E85500; E85500.
CC PIR; E90649; E90649.
CC PIR; S45229; S45229.
CC PDB; 1KY9; 03-APR-02.
CC MEROPS; S01-273; -
CC
CC SWISS-2DPAGE; P09376; COLI.
CC Ecogene; EGI0463; degp.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ_2.
CC Pfam; PF00083; trypsin_1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SM00228; PDZ; 2.
CC PROSITE; PS0106; PDZ; 2.
CC Hydrolyase; Serine protease; Heat shock; Periplasmic; Repeat; Signal;
CC Complete proteome; 3D-structure.
CC SIGNAL 1 26
CC CHAIN 27 474 PROTEASE DO.
CC DOMAIN 280 371 PDZ 1.
CC DOMAIN 377 466 PDZ 2.
CC ACT_SITE 131 131 CHARGE RELAY SYSTEM (POTENTIAL).
CC ACT_SITE 161 161 CHARGE RELAY SYSTEM (POTENTIAL).
CC ACT_SITE 236 236 CHARGE RELAY SYSTEM (POTENTIAL).
CC DISULFID 83 95
CC CONFLICT 10 10 A -> R (IN REF. 1, 7 AND 8).
CC CONFLICT 46 46 E -> Q (IN REF. 7).
CC CONFLICT 192 192 A -> G (IN REF. 1).
CC CONFLICT 467 474 STYLLMQ -> RHLPTNAVISLNPFLKTRGSPYNL (IN REF. 1).
CC SEQUENCE 474 AA; 49354 MW; 5482E596F74B6D5F CRC64;
CC
CC Query Match 17.6%; Score 317.5; DB 1; Length 474;
CC Best Local Similarity 29.8%; Pred. NO. 5.7e-11;
CC Matches 112; Conservative 51; Mismatches 108; Indels 105; Gaps 16;
CC
CC QY 15 LSVLAAVGLGLATAPAAQAPALPSQDRFA-DFFALPLDPSAMVAQAPQVNTK---- 69
CC ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
CC Db 8 LSAL-ALSLGLALSPLSATTAETSSATTAAQMPSL-----APMLEKVMPSVVSINVESTT 62
CC
CC QY 70 -----LGYNA-----VCAGTGVDPN 87
CC
CC Db 63 VNTPRMRFNQFFGDDSPFCQSGSPFQSSPFCCGGGGGNGGQGGQKFMALGSGVLDAD 122
CC
CC QY 88 -GVVLTNNHVIAGATDINAFSGVGQTYGVVDVGYDRTQDVAVLQRLGAGGLPS--AAIG 144

```

```

Db 123 KGVVTVNNHVVNDATVVKV-QISDGRKFDKWKVGDPRSDIALIQLIQNPKNLTA-KMADS 181
QY 145 GGVAVGEPVVMGNSGGGCTPRAVGRVVALGQTVQASDSUTGAE-EFLNGLIFDDAI 203
Db 182 DALRVGDYTVAGNPFGLGET-VTSGIVSALGR-----SGLNAENYENFIQTDAI 231
QY 204 QPGDSGGPVVNGLGQVGVGNTA--ASDNFQLSGGGGGAIP-----ICQAMAIAQIORS 255
Db 232 NRNSGGALVNLNGELIGINTAILAPDGGNI---GIGFAIPNNVKNLTSQWVEVGQVKR 288
QY 256 GGGSPVTHICPTAFLGLGVVDNNGN-----GARVQRVVGSSAPAAASLGISTGD 302
Db 289 G-----ELGIMGTELNSELAKAMKVDQAQRGAFVSQVLPNSSAAKAGIKAGD 334
QY 303 VITAVDGAIPNSATAM 318
Db 335 VITSLNGKPISSFAAL 350

RESULT 7
HTRA_LACIA
ID HTRA_LACIA STANDARD; PRT; 408 AA.
AC Q9LA06;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine protease do-like htra (EC 3.4.21.-) (Htra11).
GN HTRA OR IL2136.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=IL1403;
RX MEDLINE=20177820; PubMed=10712686;
RA Poquet I., Saint V., Seznec E., Simoes N., Bolotin A., Gruss A.;
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
CC -!- FUNCTION: DEGRADATES ABNORMAL EXPORTED PROTEINS. NEEDED FOR THE PRO-
CC PEPTIDE PROCESSING OF A NATURAL PRO-PROTEIN AND FOR MATURATION OF
CC A NATIVE PROTEIN. RESPONSIBLE FOR THE HOUSEKEEPING OF EXPORTED
CC PROTEINS.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (probable).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENMBL; AF155705; AAF61294.1; .
CC EMBL; AE006442; AAK06234.1; .
CC EMBL; H86891; H86891.
CC MEROPS; S01.273.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1C.

```

```

DR Pfam; PF00095; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO00834; PROTEASES2C.
DR SMART; SMO0228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Serine protease; Transmembrane; Complete proteome.
FT TRANSMEM 6
FT DOMAIN 88 284
FT DOMAIN 302 383
FT DOMAIN 127 127
FT ACT_SITE 157 157
FT ACT_SITE 239 239
FT ACT_SITE 239 239
SQ SEQUENCE 408 AA; 41648 MW; 581B90B5A7DF851 CRC64;

Query Match 17.2%; Score 310; DB 1; Length 408;
Best Local Similarity 31.3%; Pred. No. 1.3e-10;
Matches 93; Conservative 51; Mismatches 119; Indels 34; Gaps 11;

QY 79 GTGIVDPNG---VVLNNHVIAGTINAFSGGQTVGVVDVYGYRTQVAVLQIRG- 134
Db 108 GSGVIYKSGGDVAVVTVNYHVIAGNSLDVLLSG-GQVKASVVGVDYDTDLAVLKISSE 166
QY 135 -AGGLPSAIGGVAVGEPVVMGNSGGGCTPRAVGRVVALGQTVQASDSLGAETL 193
Db 167 HKVDATFADSSKLTIGEPFAVAGSPGSGPANTATETGILSATSRQVTLQE-NGQTNI 225
QY 194 NGLIQFDAIQQGDSGGPVVNGLGQVVGNTA---ASDNFQLSGGGGGAIPICQAMAIA 250
Db 226 NA-IQTDAINPGNSGGALINIEQVIGITQSKITTEDGSTVEGLGFAIPSDNVNII 284
QY 251 QQIRSGGSPVHIGTPAFLGLGVD-----NNGN-----GARVQRVVGSAAPAA 294
Db 285 NKLEADG-----KISRPA-LGIRVMDLSQLSTNDSSQKLPSSVTGGVVVSVSGSLPAA 338
QY 295 SLGISTGDTAVDGAIPNSATAMADALNHGHPGDSVSNVWQTKSGGTRTGNVTLAE 351
Db 339 SAGLKAGDVITKGVDTAVTSDTQSLYSHNINDTVKVTYY-RDGKNTADVLSK 394

RESULT 8
DEGP BRUME
ID DEGP BRUME STANDARD; PRT; 513 AA.
AC Q8VG32;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR BME11330.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mijer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA Jablonksi L., Larsen N.H., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyripides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis".
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE009571; AA152511.1; -;
 DR PIR; AD3418; AD3418.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 2.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SMO0228; PDZ; 2.
 DR PROSITE; PS0106; PDZ; 2.
 DR Hydroxylase; Serine protease; Periplasmic; Repeat; Signal;
 KW Complete proteome.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 513 PROBABLE SERINE PROTEASE DO-LIKE.
 FT DOMAIN 125 299 CATALYTIC.
 FT DOMAIN 300 391 PDZ 1.
 FT DOMAIN 414 500 PDZ 2.
 FT ACT_SITE 152 152 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 182 182 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 257 257 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 513 AA; 53514 MW; C472FEF9DFC6268 CRC64;
 Query Match 17.0%; Score 306; DB 1; Length 513;
 Best Local Similarity 28.2%; Pred. No. 2.7e-10;
 Matches 126; Conservative 43; Mismatches 142; Indels 136; Gaps 18;
 QY 1 MSNRRSLRWLLSVLAAGVGLATAPA-----QAAPPALSODRF 42
 Db 1 MSRRISNYR-----KGVAVALSAALAGAFVVTGPIGALNEARAAVHTPP--QAGF 53
 QY 43 ADFFPALPLDPSAMVAQVAPQVNNITKL----- 70
 Db 54 AD-----LVEKVRPAVSVRVKQVQETSNRGPQFFGPPGDPDLPDGHPLKRF 103
 QY 71 -----GYNVAGATGIVIDPNGVVLNNHVIAGATDINAF 106
 Db 104 DFGMEPRGDSNRNRKANKPRPHERPVAQSGFVISEDGVVVTNNHVV-----SDGDAY 160
 QY 107 SV--GSGQTGVGVYDRTQDVAVLQRCAGGLPS-----AAIG--GVVAGCEPVWAG 157
 Db 161 TVLDDTDELDAKLIGADPRTDIAVLKINA---PRKFFVYVAFGDDNKVRGDNVWVAVG 216
 QY 158 NSGQGGTTPRVPGRVVALGQTVOASDLSLTGAEETLNGLIQFDAAIQGDSGGPVVNGLG 217
 Db 217 NPFGLGGT--VTSGIVSARGRDIGAG-----PYDDFIQIDAAVKNKSGSPAFDLGG 266
 QY 218 QVVGMTAAADNQLSGGGG--FAPIGQMAIAGQIRSGG-----GSTVHLGP----- 266
 Db 267 EVIGINTAI---FSPSGGSGVGIAPFAPSSSTAKQVVDOLIKKGSVERGWTVGVIQPVTKDI 323
 QY 267 TAFILGLGVWNGNGARVQVRWGSAPASLIGITGVITAVDGAPTSATAMADALNGHH 326
 Db 324 AASLGLA---EEKGAIVASPDQDGAAGAKIGAGDVITAVNGETQDPRDLARKVANTA 379
 QY 327 PGDIVSVNWQTKSGTGTGNVTIABGP 353
 Db 380 PGEKAALTVMRKNAKEEI-NVTITAMP 405
 RESULT 9
 DEGP_BRUSU
 ID_DEGP_BRUSU STANDARD; PRT; 513 AA.
 AC Q44597;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable serine protease do-like precursor (EC 3.4.21.-).
 GN DSGP OR HTFA OR BR0611.
 OS Brucella suis, and

OS Brucella abortus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OC NCBI_TaxID=29461; 235;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP SPECIES=B.abortus; STRAIN=2308;
 RX MEDLINE=9516590; PubMed=7861951;
 RA Tatum F.M., Chevillat N.F., Morfitt D.;
 RT "Cloning, characterization and construction of htrA and htrA-like
 RT mutants of Brucella abortus and their survival in BALB/c mice."
 RL Microb. Pathog. 17:23-36(1994).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP SPECIES=B.suis; STRAIN=1330 / Biovar 1;
 RX MEDLINE=22247741; PubMed=12271122;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Seanan M.J.,
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
 RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
 RA Hoover D.L., Lindler L.B., Halling S.M., Boyle S.M., Fraser C.M.;
 RT "The Brucella suis genome reveals fundamental similarities between
 RT animal and plant pathogens and symbionts."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U07352; AA070164.1; -;
 DR EMBL; AE014368; AAN29540.1; -;
 DR PIR; I40060; I40060.
 DR TIGR; BR0611; -;
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ
 DR InterPro; IPR001254; Peptidase_S1
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 2.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SMO0228; PDZ; 2.
 DR PROSITE; PS0106; PDZ; 2.
 DR Hydroxylase; Serine protease; Periplasmic; Repeat; Signal;
 KW Complete proteome.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 513 PROBABLE SERINE PROTEASE DO-LIKE.
 FT DOMAIN 125 299 CATALYTIC.
 FT DOMAIN 300 391 PDZ 1.
 FT DOMAIN 414 500 PDZ 2.
 FT ACT_SITE 152 152 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 182 182 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 257 257 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 513 AA; 53483 MW; DE1CE91959472806 CRC64;
 Query Match 16.9%; Score 305; DB 1; Length 513;
 Best Local Similarity 28.2%; Pred. No. 3e-10;
 Matches 126; Conservative 43; Mismatches 142; Indels 136; Gaps 18;
 QY 1 MSNRRSLRWLLSVLAAGVGLATAPA-----QAAPPALSODRF 42
 Db 1 MSRRISNYR-----KGVAVALSAALAGAFVVTGPIGALNEARAAVHTPP--QAGF 53
 QY 43 ADFFPALPLDPSAMVAQVAPQVNNITKL----- 70
 Db 54 AD-----LVEKVRPAVSVRVKQVQETSNRGPQFFGPPGDPDLPDGHPLKRF 103

QY 71 -----GYNVAVGAGTGIVIDPNCVLTNNHVIAGATDINAF 106
 DB 104 DFCMEPRGDSRSDNRGKANKPRPHERFVAGSGFVISEDGYVYNNHV---SDGAY 160
 QY 107 SV--GSGQTYGVVDVYDRTQDVAVLQLRGAGCLPS-----AAIG--GGVAVGEFVAVMG 157
 DB 161 TVVLDGTELDKALIGADRTDLAVLKINA-----PKRFVYVAFGDDNKVRVGDVWVAVG 216
 QY 158 NSGGQGGTPRVPGRVVALGQTVQASDSLGTGABETLNGLIQDAAIQGDSGGPVVNGLG 217
 DB 217 NPFGLGGT--VTSGVISARGRDIGAG-----PYDDFIQIDAAVKNKSGGPAFDLSG 266
 QY 218 QVGVNTASDNFQLSGQGG--FAIPICQAMAIAGQIRSGG-----GSPVTHVIG----- 266
 DB 267 EVIGINTAI--FSPSGSGVGTAFALPSTAKQVVDQLIKKSGVERGMVGVQIQVTKDI 323
 QY 267 TAFPLGJVVDNNGNGARVQVVGSPAASLGISTGVTITAVDGPINSATAMADALNGHH 326
 DB 324 AASLGLA---BEKGAIVASPDQDGAAGKAGDVITAVNGETVQDPRDLARKVANIA 379
 QY 327 PGDIVSNVQTKSGGRTGNVTLAECF 353
 DB 380 PGEKAALTVMRKNKABEI-NVTIAAMP 405

RESULT 10
 HTRA_BACSU
 ID HTRA_BACSU STANDARD; PRT; 449 AA.
 AC O34358;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable serine protease do-like htra (EC 3.4.21.-).
 GN HTRA_OR BSU12900.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 CX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Devine K.M.;
 RT "Sequence of the Bacillus subtilis genome between xlyA and ykoR";
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=38044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 Lee S.M., Levine A., Liu H., Masuda S., Mael C., Medigue C.,
 Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 Parro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,
 Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 Sekiguchi J., Sekowska A., Seror S.J., Serrif P., Shin B.S., Soldo B.,
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 Takeuchi M., Tamakoshi A., Tanaka T., Terpetra P., Tognoni A.,
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,

Viari A., Wambutt R., Wedler E., Wedler H., Weitzneger T.,
 Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
 "the complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis."; 249-256 (1997).
 [3]
 RN Nature 390:249-256 (1997).
 RP TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX MEDLINE=20158875; PubMed=10692364;
 RA Noone D., Howell A., Devine K.M.;
 RT "Expression of ykda, encoding a Bacillus subtilis homologue of Htra,
 is heat shock inducible and negatively autoregulated";
 RL J. Bacteriol. 182:1592-1599 (2000).
 [4]
 RN TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX MEDLINE=20576168; PubMed=11133960;
 RA Noone D., Howell A., Collier R., Devine K.M.;
 RT "Ykda and yvta, Htra-like serine proteases in Bacillus subtilis,
 engage in negative autoregulation and reciprocal cross-regulation of
 ykda and yvta gene expression";
 RL J. Bacteriol. 183:654-663 (2001).
 [5]
 RN TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX MEDLINE=21439741; PubMed=1155295;
 RA Hyttilaenen H.-L., Bolhuis A., Darmon E., Muukkonen L., Koski P.,
 Vitikainen M., Sarvas M., Pragai Z., Bron S., van Dijk J.M.,
 Kontinen V.P.;
 RT "A novel two-component regulatory system in Bacillus subtilis for the
 survival of severe secretion stress";
 RL Mol. Microbiol. 41:1159-1172 (2001).
 CC -!- FUNCTION: May be involved in processing, maturation, or secretion
 of extracellular enzymes.
 CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
 CC -!- INDUCTION: Transcription is css dependent. Induced by heat shock
 during exponential growth and by heterologous amyloses at the
 transition phase of the growth cycle. Negatively regulates its own
 expression during exponential growth and during heat shock.
 CC -!- MISCELLANEOUS: Inactivation results in compensating overexpression
 of yvta, especially during stress conditions.
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL; AJ002571; CA05570.1; -;
 DR EMBL; Z91110; CAB13147.1; -;
 DR PIR; A69643; A69643.
 DR MEROPS; S01.273; -;
 DR Subtilist; BG12608; htra.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR InterPro; PF00595; PDZ; 1.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00226; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 DR HydroLase; Protease; Serine protease; Heat shock; Transmembrane;
 Complete proteome.
 KW DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 45 67 POTENTIAL.
 FT DOMAIN 68 49 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 103 108 POLY-SER.
 FT

FT DOMAIN 146 152 POLY-SER.
 FT DOMAIN 348 437 PDZ
 FT ACT SITE 179 179 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 209 209 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 230 290 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 449 AA; 47713 NW; E12B07A9018BE414 CRC64;

Query Match 16.4%; Score 295.5; DB 1; Length 449;
 Best Local Similarity 29.3%; Pred No. 8.9e-10;
 Matches 93; Conservative 53; Mismatches 126; Indels 45; Gaps 11;

QY 53 SAMVAQAPQVNNINKLGYNNA-----VGAGTGIV-IPNG--VVLITNN 94
 DB 119 SDVDELSPAIVGITNLQAQNSSLFGSSSDSDSDTESGSGGVIFKKGKAYIITNN 178
 QY 95 HVIAGATDINAFVSGGQTYGVVDVDRDQVAVLQRLGAGLPSAIG--GGVAVGEP 152
 DB 179 HVVEGASLKV-SLYDGTETVAKLVGSDSLTDLAVLQISDDHVTKVANFGDSSDLRTGET 237
 QY 153 VVAMNNGSGGGGTTPRAVGRYVALGQTYQASDSLTGAETLGLIQFDAATQPGDGGPV 212
 DB 238 VIAIGDPLGKLSRTVTCQIVSGVDRTVMS---TSAGETSLNVIQTDALINFGNSGGL 294
 QY 213 VNGIGQVVVGMNT--AASDNFOLSGGGQFAIPICQAMAIAGQIRSGGSPVHIGTAF 270
 DB 295 LNTDGKIVGINSKISDDVE---GIGFAIPSNVDRKPIABELLSKQIERPVIG-VSML 349
 QY 271 GLGVVDNN-----GNGARVQVVGSAPAAASLGTSTGDTVITAVDGAIPINSATA 317
 DB 350 DLEQVPQYQEGTGLFGSQLNKVYIREVASSGPAEKAGIKADIIILGKKEIDTGE 409
 QY 318 MADAL-NGHHPGDVISV 333
 DB 410 LRNLYKDAKIGDTVEV 426

RESULT 11
 ID HTOA_HAEIN STANDARD; PRT; 466 AA.
 AC P45129;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable periplasmic serine protease dc/hhoA-like precursor
 DE (EC 3.4.21.-).
 GN H11259.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A., Bult C.J., Tomb J.-P., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Burton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd.",
 RL Science 269:496-512(1995).
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -!- SIMILARITY: SEEMS TO BE AN INTERMEDIATE FORMS BETWEEN E.COLI HTRA
 CC (PROTEASE DO) AND HHOA.
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

ENBL; U32805; AAC22906.1; -;
 DR FIR; A64113; A64113.
 DR MEROPS; S01.274; -;
 DR TIGR; H11259; -;
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR001478; PDZ_1.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 2.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; SM00228; PDZ; 2.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS0106; PDZ; 2.
 DR Hydrolase; Serine protease; Periplasmic; Repeat; Signal;
 KW Complete proteome.
 KW SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 466 PROBABLE PERIPLASMIC SERINE PROTEASE
 FT DOMAIN 270 361 DO/HHOA-LIKE.
 FT ACT SITE 120 120 PDZ 1.
 FT ACT SITE 150 150 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 226 226 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 466 AA; 49434 NW; ED050A00047B5851 CRC64;

Query Match 16.4%; Score 295; DB 1; Length 466;
 Best Local Similarity 28.1%; Pred No. 9.9e-10;
 Matches 108; Conservative 54; Mismatches 124; Indels 98; Gaps 14;

QY 17 VLAAVGLGLATAP-----AQAAPPALSDQRFADFPALPLDPSAMVAQVAPVNNITK 69
 DB 7 VLNSIALGLSVLSTSFVAHVAAQATLPFSVSNQSLAP-----MLEKQPAVVVTLSE 58
 QY 70 -----LGNNAVAGAGTGIVDPN-GVVLTN 93
 DB 59 GKAKVDSRSPFLDDIPEEFKFFGDRFAEQPGGSGSKRNFRLGSGVVIINASKGYVLTN 118
 QY 94 NHVIAGATDINAFVSGGQTYGVVDVDRDQVAVLQRLGAGLPSA--AIGGVAVGE 151
 DB 119 NHVIDGADKIV-QLDGREGKAKLVGKDEQSDIALVQLEKPSNLTIKFADSKLAVGD 177
 QY 152 PVVAMNNGSGGGTTPRAVGRVVALGQTYQASDSLTGAEB-TLNGLIQFADAAIQPGDGG 210
 DB 178 FTVAIGNPFLGQGT--VTSGIVSALGRS-----TGSDSGTENYIQTDAAVNRNGSG 228
 QY 211 PVVAMNGLGVGMNTAASDNFOLSGGGQ--FAIPGQAMAIA-----GOIRSGGSPV 262
 DB 229 ALVNLGELIGINTAI---ISPSGNGAGIAFAIPSNQASNLVQQLLEFGQVRRG----- 279
 QY 263 HIGTAFGLGVVDNNGN-----GARVORVVGSAAPASLGISTGDTITAVDGP 311
 DB 280 -----LLGKGGELNADLAKAFNVSAQQGAFVSEVLPSKAAEAKGLKAGDIITAMNGQK 333
 QY 312 INSATAMADALNGHHPGDVISVNW 335
 DB 334 ISSPAEIRAKIATGAGKEISLTY 357

RESULT 12
 ID DEGP_RHIME STANDARD; PRT; 504 AA.
 AC Q52854;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable serine protease do-like precursor (EC 3.4.21.-).

DEGP1 OR DEGP OR R01021 OR SMC02365.
Rhizobium meliloti (Sinorhizobium meliloti).
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
NCBI_TaxID=382;
[1]
SEQUENCE FROM N.A.
STRAIN=1021;
MEDLINE=96146524; PubMed=8550509;
Glazebrook J., Ichige A., Walker G.C.;
RT "Genetic analysis of Rhizobium meliloti bacA-phoA fusion results in
RT identification of degp: two loci required for symbiosis are closely
RT linked to degp.";
RL J. Bacteriol. 178:745-752 (1996).
[2]
SEQUENCE FROM N.A.
STRAIN=1021;
MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boissard P., Becker A., Soutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kies E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puchler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882 (2001).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U31512; AAC43669.1; ALT_INIT.
CC EMBL; AL591785; CAC45593.1; -.
CC MEROPS; S01.273; -.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ; 2.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SM00228; PDZ; 2.
CC PROSITE; PS0106; PDZ; 2.
CC Hydrolase; Serine protease; Periplasmic; Repeat; Signal;
CC Complete proteome.
CC SIGNAL 1 26
CC CHAIN 27 504
CC DOMAIN 113 286
CC DOMAIN 287 378
CC DOMAIN 401 491
CC ACT_SITE 140 140
CC ACT_SITE 170 170
CC ACT_SITE 244 244
CC CONFLICT 14 15
CC CONFLICT 39 147
CC
CC CONFLICT 464 504
CC PSFANVVDVAVSVVVRVQAREVSDENFTDFGGRGF
CC EDLPEDHPLRRFFEFAPRDRDRDRGRGRLRP
CC RAQGGFFITDGYLVNHNHVSQSA -> AVSPMWTFF
CC RRRSSPSACRHVNATMKATSPSIAAAGSRTCKRTICG
CC VSSANSRLVKMTPIVGTAAVRVAKVSVGRKAPASSP
CC KIVTSSPTTSSPART (IN REF. 1)
CC KSAADVLLKVNNAKGRSKALFQEAQGRFVALPITOG
CC -> NRQTTFR (IN REF. 1).
CC
CC SEQUENCE 504 AA; 53035 MW; D7E82BB981EA23C CRC64;
Query Match 16.1%; Score 291; DB 1; Length 504;
Best Local Similarity 33.3%; Pred. No. 1.8e-09;
Matches 97; Conservative 39; Mismatches 117; Indels 38; Gaps 12;
QY 79 GTGIVIDNGVVLNNHVIAGATINAFSV--GSGQYGVVVGVYDRTQVAVLQL--RG 134
DB 124 GSGFFITDGYLVNHNHVV---SDGSAFTVIMNDGTDLAKLVGKDSRTDLAVLKVDDKR 180
QY 135 AGGLPSAIGGVAVGEPVAVMGSGGQGPRAVPRVAVGALGOTVQVQASDLSLTGAETLN 194
DB 181 KFTYVSFADDEKVRVGVVAVGPPFLGGTVTA--GIISARGDGGSG-----PYD 230
QY 195 GLIIFDAAIQDSDGGPVVNGLVGVVGMNTAASNFOLSQGGQ--PAIPGQAWAIAGQ 252
DB 231 DYLQVDAAVNRGSGGPTFNLSGEVVGINTAI---FSPSGGVNVTAFAPASVAKDVVDS 287
QY 253 IRSGG-----GSFTVHIGPTA-----FLGLGVVDNNGCARVQRVVGSAAPASLGISTGDV 303
DB 288 LKDGTVSRGVLGVQIQVTKDIAESLGL----SEANGALVVEPQAGSPGSKAGIKNGDV 343
QY 304 ITAVDGAIPNGATAMADALNGHPDGVISVN-WOTKSGTGTGNVTLAEGP 353
DB 344 VTALNGEVPKPRDLARRVAAALPGSTAEVTLW--RSKSEIVNLEIGTLP 392
RESULT 13
DEGP BARHE
ID DEGP BARHE STANDARD; PRT; 503 AA.
AC P54925;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable periplasmic serine protease DO-like precursor (EC 3.4.21.-)
DE (Antigen htra).
GN DEGP OR HTRA.
OS Bartonella henselae (Rochalimaea henselae).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bartonellaceae; Bartonella.
OX NCBI_TaxID=38323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 49882 / Houston 1;
RX MEDLINE=94299828; PubMed=8027347;
RA Anderson B., Sims K., Regnery R., Robinson L., Schmidt M.J.,
RA Goral S., Hager C., Edwards K.;
RT "Detection of Rochalimaea henselae DNA in specimens from cat scratch
RT disease patients by PCR";
RL J. Clin. Microbiol. 32:942-948 (1994).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L20127; AAA97430.1; -.
CC MEROPS; S01.273; -.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ; 2.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SM00228; PDZ; 2.
CC PROSITE; PS0106; PDZ; 2.
CC Hydrolase; Serine protease; Periplasmic; Repeat; Signal.
CC POTENTIAL.
CC SIGNAL 1 18
CC CHAIN 19 503
CC PROBABLE PERIPLASMIC SERINE PROTEASE DO-
CC LIKE.

DOMAIN 286 357 PDZ 1.
 FT ACT SITE 419 466 PDZ 2.
 FT ACT SITE 143 173 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 173 247 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 503 AA; 54114 MW; 6CD9F474328AF9E CRC64;

Query Match
 Best Local Similarity 16.1%; Score 290; DB 1; Length 503;
 Matches 113; Conservative 42; Mismatches 135; Indels 118; Gaps 16;

QY 22 GLGLATAFAQA---APPALSDRFPALPDLPSAMVAQVAPQVNTKLG----- 71
 DB 29 GSSLMTTKAHANSVSSLMQOQFAD-----IVSQVPAVSVQVKNKKKEWF 78
 QY 72 -----YN-----NAVAGAGTIVIDPQGVV 90
 DB 79 FSDFFSTPGFDLPDQHLKFFQDFYNRDKPSNKSLSQSRHLRPIAFGGFFISSDGI 138
 QY 91 LTNHVIAGATINAFSVSGQTVGVVVDRTQDVAVLQLRGAGGLPSAAIG--CGVA 148
 DB 139 VTNNHVISDGTSY-AVLLDDGTETLNAKLIGTDPRDLAVLKNKRFYSYVDFGDDSKLR 197
 QY 149 VGEFVWAMNSGGGGTPRAVPGRVWALGOTVQASDSLTGAETLNGLIQFDAAIQGDS 208
 DB 198 VGDVVAIGNFPLGGTUTA--GIVSARGR-----DIGTG---VYDFIQLDAVNRGNS 247
 QY 209 GGPVVGNGVGVGNTAASDNFQSQGGQ--FAIPICQAMAIAGQIRSGGSPVTHIGP 266
 DB 248 GGPTFDLNGKVVGVNTAI--FSPSGGVGIAFAIPAATANEVQQL----- 291
 QY 267 TAPLGLGVVNGNGARQVQV-----VG-----SAPASIGISTGDVITA 306
 DB 292 ---TEKGLVQSGWGLVQIQPTKEISDLSIGLKEAKGALITDPLKGPAAKAGIKAGDVIIS 348
 QY 307 VDGAPINSATAMADALNGHPGDVIVSN-WQTSKGGTRTGNVTLAEGP 353
 DB 349 VNGEKINDVRDLAKRIANMSGPEVTLGVW--KSGKEENIKVKLDSNP 394

RESULT 14

DEGI ARATH
 ID DEGI ARATH STANDARD; PRT; 437 AA.
 AC 022609; 09LK85;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Protease Do-like 1, chloroplast precursor (EC 3.4.21.-).
 GN DEGI1 OR DEGP OR AT3G27925 OR K16N12.18.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OK NCBI_TaxID=3702;
 RN [1]
 RN SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RP MEDLINE=98175982; PubMed=9507020;
 RX Itzhaki H., Naveh L., Lindahl M., Cook M., Adam Z.;
 RA "Identification and characterization of Degp, a serine protease
 RT associated with the luminal side of the thylakoid membrane.";
 RL J. Biol. Chem. 273:7094-7098 (1998).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RC MEDLINE=20363099; PubMed=10907853;
 RA Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
 RL and BAC clones.";
 RN DNA Res. 7:217-221 (2000).
 RN [3]
 RN SEQUENCE OF 104-118.
 RP STRAIN=cv. Columbia;
 RC

RA Kieselbach T., Bystedt M., Schroeder W.P.;
 RL Submitted (JUL-2000) to Swiss-Prot.
 CC -I- FUNCTION: Serine protease that is required at high temperature.
 CC May be involved in the degradation of damaged proteins. In vivo,
 CC can degrade beta-casein.
 CC -I- ENZYME REGULATION: Inhibited by phenylmethylsulfonyl fluoride and
 CC O-phenanthroline.
 CC -I- SUBCELLULAR LOCATION: BOUND TO LUMINAL SIDE OF THE THYLAKOID
 CC MEMBRANE.
 CC -I- INDUCTION: By heat shock.
 CC -I- SIMILARITY: Belongs to peptidase family S2C.
 CC -I- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC EMBL; AF028842; AAC39436.1; -;
 DR EMBL; AF000371; BAB02539.1; -;
 DR EMBL; AF001302; BAB02539.1; JOINED.
 DR MEROPS; S01.279; -;
 DR InterPro; IPR009003; Cys Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 DR Hydrolase; Serine protease; Transit peptide; Chloroplast; Thylakoid.
 KW TRANSIT ?
 FT TRANSIT ? 103
 FT CHAIN 104 437
 FT DOMAIN 152 321
 FT DOMAIN 324 421
 FT ACT SITE 171 171
 FT ACT SITE 201 201
 FT ACT SITE 280 280
 FT CONFLICT 12 23
 FT CONFLICT 36 36
 FT CONFLICT 54 54
 FT CONFLICT 60 60
 FT CONFLICT 64 64
 FT CONFLICT 68 69
 FT CONFLICT 355 355
 FT CONFLICT 381 381
 FT CONFLICT 416 416
 FT CONFLICT 437 AA; 46213 MW; 1497B1AB3F5FF2A4 CRC64;
 SQ SEQUENCE 437 AA; 46213 MW; 1497B1AB3F5FF2A4 CRC64;

Query Match
 Best Local Similarity 16.1%; Score 289.5; DB 1; Length 437;
 Matches 111; Conservative 56; Mismatches 143; Indels 81; Gaps 16;

QY 7 RSLKRWMLLSVLA-----VGLGLATAFAQAAPP-----ALSQDRFADPPA---- 47
 DB 42 RSKRYFRILSKPLPLNDNNGDDGDTLLITPFSVAVKPFLLCTSVLSFSLFAASPAVESA 101
 QY 48 -----LPLDPSAMV---AQVAPQVNTKLVNNAV-----GAGTGVID 85
 DB 102 SAFVSVTPKQLQOTDELATVRLFQNTFSVVITNLAVRQDAFTLDVLEVFQSGSGGFWD 161
 QY 86 PNGVVLTNHVIAGATINAFSVSGQTVGVVVDRTQDVAVLQLRGAGG-----LPSA 141
 DB 162 KQHIVTNYHVIRGASDLRV--TLADQTTFDKVVGFDDKDAVLRIDAPKNLRPIP-V 219
 QY 142 AIGGVAVGEPVWAMNSGGGGTTPRAVPGRVWALGOTVQASDSLTGAETLNGLIQFDA 201

Db 220 GVADLLVGKVFAGNPFGLDHT--LTTGVISGLRREI--SSAATG--RPTQDVLOTDA 273
QY 202 AIQPGDSGGVWAGLGQVGMNTAASDNFQSQGGQGFAPICQAMAIAGQIRSGGSPT 261
Db 274 AINPNSGGPGLDSSGFLICGINTAIYSPGAS--SGVFSIPVTVGGIVDQL----- 324
QY 262 VHGPTAFGLGV-----VDNNG-NGARVQVVSAPAAASIGIST-----GD 302
Db 325 VRFGKVRTPILGKFAFDQVEQLGVSGLLLDAPPSPGAGKAGLQSTKRDGVRILIGD 384
QY 303 VITAVDGAIPNSATAMADALNHHHPGDVISV 333
Db 385 IITSVNGTKVNSGSDLYRLDCKQGVDEVTV 415

RESULT 15
DEGP CHLTR STANDARD; PRT; 497 AA.
AC P18584; O84830;
DT 01-NOV-1990 (Rel. 16, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-) (59 kDa
DE immunogenic protein) (SK59).
GN DEGP OR HTRA OR CT823.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
CX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serovar L2;
RX MEDLINE=90337348; PubMed=2379836;
RA Kahane S., Weinstein Y., Sarov I.;
RT "Cloning, characterization and sequence of a novel 59-kDa protein of
RT Chlamydia trachomatis";
RL Gene 90:61-67(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/Cx;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis";
RL Science 282:754-759(1998).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC -!- CAUTION: REF.1 SEQUENCE WAS INCORRECT: DUE TO SEQUENCING ERRORS
CC THE AUTHORS TRANSLATED THEIR PUTATIVE 59 kDa IMMUNOGENIC PROTEIN
CC ON THE COMPLEMENTARY STRAND TO THAT OF WHAT SEEMS TO BE THE REAL
CC PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AR001355; AAC68420.1; -
CC EMBL; M31119; AAA23116.1; -
CC PIR; H71465; H71465.
CC PIR; H71465; P18584; -
CC HCI-2DPAGE; P18584; -
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ_
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR008256; Peptidase_S1B_Va.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ; 2.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00834; PROTEASES2C.

PRINTS; PR00839; V8PROTEASE.
SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
KW Hydrolase; Serine protease; Repeat; Signal; Antigen;
KW Complete proteome.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 497 PROBABLE SERINE PROTEASE DO-LIKE.
FT DOMAIN 128 289 CATALYTIC.
FT DOMAIN 290 381 PDZ 1.
FT DOMAIN 394 485 PDZ 2.
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 497 AA; 53244 MW; 86A5E31BB84A38BA CRC64;

Query Match 16.0%; Score 288; DB 1; Length 497;
Best Local Similarity 32.3%; Pred. No. 2.6e-09;
Matches 95; Conservative 46; Mismatches 109; Indels 44; Gaps 12;
QY 79 GTGIVDPNGVLTNNHVIAGATDINAFSVSGSGOTYGVVDYDRTQDVAVLQLRGAGL 138
Db 127 GTGFIIVSEGGVVTNNHVVEDAGKIHV--TLHDGQYTKIVGLDPTDLAVIKIQ-AEKL 184
QY 139 PSAAIIGG--VAVGEPVVMGNSGGGGTTPRAVGRVVALGQT----VOASDSLTAET 192
Db 185 PELTFGNSDQLIGDWAIAGNPFGLQAT--VTGVISAKGRNQLHIVDFED----- 234
QY 193 LNLGIQFDAATQPCDSGSPVNGLVGMNTAASDNFQLSQG--GQGFAPIGQAMAIA 250
Db 235 ---FIQTDAAINPNSGGPGLNINQVIGVNTAIVSG---SGYIGIGFAPISLMKRV 288
QY 251 GQIRSGGSPVTHIGTAFGLGVVDN-----NNGARVQVVSAPAAASIGISTG 301
Db 289 DGLISDGQVTRGFLGT---LQPIDSELATCYKLEKYVYGVALTVDVVKGSFAEKAGLQ 344
QY 302 DVITAVDGAIPNSATAMADALNHHHPGD--VTSVNWQTKSGGTRTGNTVLAEGP 353
Db 345 DVIVAYNGKEVESLSALRNATSLMMPGTRVVLKI---VREGKTIPIPTVTQIP 395

Search completed: June 22, 2004, 17:16:46
Job time : 7.99901 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:55:54 ; Search time 29.0268 Seconds
(without alignments)

3858.816 Million cell updates/sec

Title: US-09-886-349A-2

Perfect score: 1802

Sequence: 1 MNSRRRSRLRWLSVLAA.....QTKSGGTRGNVTLAEGPPA 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeop.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1792	99.4	355	16 Q07175	Q07175 mycobacteri
2	1792	99.4	355	16 Q70289	Q70289 mycobacteri
3	1306	72.5	361	2 Q50320	Q50320 mycobacteri
4	1272.5	70.6	354	16 Q9CCY9	Q9CCY9 mycobacteri
5	469.5	26.1	446	16 Q8VKA4	Q8VKA4 mycobacteri
6	469.5	26.1	464	16 Q7U0X2	Q7U0X2 mycobacteri
7	468.5	26.0	464	16 Q53896	Q53896 mycobacteri
8	460	25.5	382	16 Q9CD67	Q9CD67 mycobacteri
9	460	25.5	452	2 Q9Z5G6	Q9Z5G6 mycobacteri
10	407	22.6	542	16 Q9F8K9	Q9F8K9 streptomyc
11	396	22.0	519	16 Q93J30	Q93J30 streptomyc
12	391	21.7	375	16 Q83G87	Q83G87 synecococc
13	382	21.2	394	16 P72780	P72780 synecocyst
14	382	21.2	472	16 Q82EM9	Q82EM9 streptomyc
15	382	21.2	619	16 Q82IL8	Q82IL8 streptomyc
16	379.5	21.1	473	16 Q8FR17	Q8FR17 corynebacte

17	373.5	20.7	500	2	Q9KJN6	Q9KJN6 myxococcus
18	372	20.6	407	16	Q8YTF9	Q8YTF9 anabaena sp
19	366	20.3	525	16	Q89QJ8	Q89QJ8 bradyrhizob
20	364.5	20.2	675	15	Q8G6T3	Q8G6T3 bifidobacte
21	363	20.1	395	15	Q7V5C8	Q7V5C8 prochloroco
22	362	20.1	525	16	Q8PMV4	Q8PMV4 xanthomonas
23	358	19.9	413	16	Q8NS10	Q8NS10 corynebacte
24	355.5	19.7	452	16	P73354	P73354 synecocyst
25	354.5	19.7	490	16	Q8XPT5	Q8XPT5 raietonia s
26	354	19.6	371	16	Q31388	Q31388 bradyrhizob
27	354	19.6	433	16	Q97GD5	Q97GD5 clostridium
28	353	19.6	326	16	Q8DMV9	Q8DMV9 synecococc
29	349	19.4	447	16	Q8R756	Q8R756 thermocanae
30	347.5	19.3	391	16	Q89915	Q89915 clostridium
31	347.5	19.3	432	16	Q7U8K9	Q7U8K9 synecococc
32	347	19.3	374	16	Q7U495	Q7U495 synecococc
33	347	19.3	465	16	Q920E6	Q920E6 rhizobium m
34	347	19.3	525	16	Q8PB56	Q8PB56 xanthomonas
35	346	19.2	389	16	Q8DL28	Q8DL28 synecococc
36	345	19.1	339	16	Q89RP2	Q89RP2 bradyrhizob
37	345	19.1	362	16	Q9S2R5	Q9S2R5 streptomyc
38	344	19.1	511	2	Q8RTX2	Q8RTX2 xanthomonas
39	343	19.0	457	16	Q8ZB58	Q8ZB58 versinia pe
40	343	19.0	463	16	Q8D1R0	Q8D1R0 versinia pe
41	342.5	19.0	528	16	Q89S21	Q89S21 bradyrhizob
42	341.5	19.0	514	16	Q9PBA3	Q9PBA3 xylella fas
43	340.5	18.9	505	16	Q8Y016	Q8Y016 raietonia s
44	338.5	18.8	629	16	Q7UXF4	Q7UXF4 rhodopirell
45	335.5	18.6	429	16	Q8Y7Z0	Q8Y7Z0 anabaena sp

ALIGNMENTS

RESULT 1

007175 PRELIMINARY: PRT; 355 AA.

ID Q07175
AC Q07175;
DT 01-JUL-1997 (TRENBLrel. 04, Created)
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein (Serine protease, putative).
GN PEPA OR RV0125 OR MTCI418B.07 OR MT0133.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (APR-2001) to the ENBL/GenBank/DBJ databases.

```

CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; Z96071; CAB09453.1; -.
DR EMBL; AE006925; AAK44357.1; -.
DR PIR; F70983; F70983.
DR TIGR; MT0133; -.
DR TubercList; rv0125; -.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ_
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR008256; Peptidase_S1B_v8.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Hypothetical protein; Serine protease; Protease;
KW Complete proteome.
KW Complete proteome.
SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF92 CRC64;

Query Match 99.4%; Score 1792; DB 16; Length 355;
Best Local Similarity 99.4%; Pred. No. 3.5e-87;
Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQV 60
QY 61 PQVNVNINIKLGYNNVAGAGTGIVDPNGVVLNNHVIAGATDINAFSVSGGQTYGVVDVVG 120
DB 61 PQVNVNINIKLGYNNVAGAGTGIVDPNGVVLNNHVIAGATDINAFSVSGGQTYGVVDVVG 120
QY 121 YDRTQDVAVLQIRGAGGLPSAAGGGVAVGEPVVMGNSGGGQGTTPRAVGRVVALGQTV 180
DB 121 YDRTQDVAVLQIRGAGGLPSAAGGGVAVGEPVVMGNSGGGQGTTPRAVGRVVALGQTV 180
QY 181 QASDLSLTGAEETLNLQIFDAAIQPGDSGGPVVNGLGQVVGMMNTAASDNFQLSQGGQGFA 240
DB 181 QASDLSLTGAEETLNLQIFDAAIQPGDSGGPVVNGLGQVVGMMNTAASDNFQLSQGGQGFA 240
QY 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAAPAASLGIST 300
DB 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAAPAASLGIST 300
QY 301 GDVITAVDGAIPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGAIPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 3
Q50320 PRELIMINARY; PRT; 361 AA.
ID Q50320;
AC Q50320;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 34kDa protein precursor.
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JD88/107;
RX MEDLINE=95005449; PubMed=7921248;
RA Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
RT "Identification and characterisation of a putative serine protease
expressed in vivo by Mycobacterium avium subsp paratuberculosis.";
RL Microbiology 140:1977-1982(1994).
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC EMBL; Z23092; CAA80638.1; -.
DR PIR; S47170; S47170.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001540; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.

Query Match 99.4%; Score 1792; DB 16; Length 355;
Best Local Similarity 99.4%; Pred. No. 3.5e-87;
Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWSLLSVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQV 60
QY 61 PQVNVNINIKLGYNNVAGAGTGIVDPNGVVLNNHVIAGATDINAFSVSGGQTYGVVDVVG 120
DB 61 PQVNVNINIKLGYNNVAGAGTGIVDPNGVVLNNHVIAGATDINAFSVSGGQTYGVVDVVG 120
QY 121 YDRTQDVAVLQIRGAGGLPSAAGGGVAVGEPVVMGNSGGGQGTTPRAVGRVVALGQTV 180
DB 121 YDRTQDVAVLQIRGAGGLPSAAGGGVAVGEPVVMGNSGGGQGTTPRAVGRVVALGQTV 180
QY 181 QASDLSLTGAEETLNLQIFDAAIQPGDSGGPVVNGLGQVVGMMNTAASDNFQLSQGGQGFA 240
DB 181 QASDLSLTGAEETLNLQIFDAAIQPGDSGGPVVNGLGQVVGMMNTAASDNFQLSQGGQGFA 240
QY 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAAPAASLGIST 300
DB 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAAPAASLGIST 300
QY 301 GDVITAVDGAIPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGAIPINSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 2
Q7U2S9 PRELIMINARY; PRT; 355 AA.
ID Q7U2S9;
AC Q7U2S9;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable serine protease PEPA (EC 3.4.21.-).
DE PEPA OR MB0130.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Biglimer K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,

```



```
KW Hydrolase; Protease; Serine protease; Signal.
FT SIGNAL 1 38 POTENTIAL
SQ SEQUENCE 361 AA; 35709 MW; 30FEF78DF3C411 CRC64;

Query Match 72.5%; Score 1306; DB 2; Length 361;
Best Local Similarity 70.4%; Pred. No. 1.6e-61;
Matches 255; Conservative 41; Mismatches 58; Indels 8; Gaps 3;

QY 1 MNSRRLRSRLSWLLSVLAAYVGLG-----ATAPQAAPALSDQRFADFPALPLDPS 53
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 53
1 MSKSHHRSVWMSLVGVTVVGLGSLGSLGVLGSLGSLGSLGSLGSLGSLGSLGSLGSLG 60
QY 54 AMYAQAPOVNIINTKLYNNVAGCTGIVDPNGVLTNNHVIAGATDINAFVSGSGQT 113
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 113
61 AMVGVGPQWNIDTFKYNVAGCTGIVDPNGVLTNNHVIAGATDINAFVSGSGQT 120
QY 114 YGVVVYGVYDRTQDAVQLRGAGLPSAAIGGVAVGEPVWAMGNSGGCGTTPRAVGRV 173
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 173
121 YAVDVYGVYDRTQDAVQLRGAGLPSAAIGGVAVGEPVWAMGNSGGCGTTPRAVGRV 180
QY 174 VALGQTVQASDSLTAETLGLIQDAAIQPDGSGPVVNGLGQVGMNTAASDNFOLS 233
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 233
181 VALNQSVSNTDLTGAEQLGLIQADAPIKDGSGGPNWNSAGQVIGVDTAATSYKMS 240
QY 234 QGQGGFAIPIGQAMAIAGQIRSGGGSPVHIGPTAFGLGVVDNNGNGARVQVVGSA 293
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 293
241 -GGQGGFAIPIGRAMAVANQIRSGAGSNTVHIGPTAFGLGVVDNNGNGARVQVVGSA 299
QY 294 ASLIGSTGVITAVDGAIPNSATAMADALNGHHPGDVSNVWQTSKGGTRTGNVLA 353
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 353
300 AAGTAFAGVITGVDTVPINGATSMTEVLPVPHPGDTIAVHFRSVDGGERTANTILAS 359
QY 354 PA 355
DB |||
360 PA 361

RESULT 4
Q9CCY9 PRELIMINARY; PRT; 354 AA.
AC Q9CCY9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable secreted serine protease.
GN ML2659.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Bauman D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagers K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL583926; CAC32191.1; -.
DR PIR; A87242; A87242.
DR Leptoma; ML2659; -.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
```

```
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR002554; Peptidase_S1.
DR InterPro; IPR008254; Peptidase_S1B_V8.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00889; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 354 AA; 35265 MW; 612F23261BC9EA4A CRC64;

Query Match 70.6%; Score 1272.5; DB 16; Length 354;
Best Local Similarity 70.1%; Pred. No. 9e-60;
Matches 251; Conservative 42; Mismatches 58; Indels 7; Gaps 3;

QY 1 MNSRRSLRSRLSWLLSVLAAYVGLGLATAPQA---APPALSDQRFADFPALPLDPSAMVA 57
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 57
1 MSRQPHRSRLSWLVSLLAALGLSLAVVPGSATSPSTLALDRFSNRPPPLPLNPAAM-- 58
QY 58 QVAPQVNIINTKLYNNVAGCTGIVDPNGVLTNNHVIAGATDINAFVSGSGQTGYVD 117
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 117
59 -VAFQVNIISRLGYNVAGAGTGIVIDSSGVLTNNHVISGATDLSAFDVGNGKTYGVD 117
QY 118 VVGVDRTQDAVQLRGAGLPSAAIGGVAVGEPVWAMGNSGGCGTTPRAVGRVWALG 177
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 177
118 VVGVDRTQDAVQLRGAGLPSAAIGGVAVGEPVWAMGNSGGCGTTPRAVGRVWALG 177
QY 178 QTVQASDSLTAETLGLIQDAAIQPDGSGGVVNGLGQVGMNTAASDNFOLSQGG 237
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 237
178 QTVQASDELTGAQETLSGLIQDAIPIKPDGSGGVVNSRGQVGMNTAATDNYKM-LGGQ 236
QY 238 GFAIPIGQAMAIAGQIRSGGGSPVHIGPTAFGLGVVDNNGNGARVQVVGSAASLG 297
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 297
237 GFAIPIGQAMAEVVGAIIRSGAGSNTVHIGPTAFGLGVVDNNGNGARVQVVGSAASLG 296
QY 298 ISTGDVITAVDGAIPNSATAMADALNGHHPGDVSNVWQTSKGGTRTGNVLAEGPPA 355
DB |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 355
297 ISVGDITISVDGVFISEATAMTNVLPVPHPGETVAVNYSAGGDLTANVTLASGPPA 354

RESULT 5
Q8VKA4 PRELIMINARY; PRT; 446 AA.
AC Q8VKA4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Heat shock protein HtrA, putative.
GN MT1011.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleisichman R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA DeCher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE006985; AAK45259.1; -.
DR TIGR; MT1011; -.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
```

```

DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR SMART; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
SQ SEQUENCE 446 AA; 44484 MW; 54170CBEA8E872B CRC64;

Query Match 26.1%; Score 469.5; DB 16; Length 446;
Best Local Similarity 36.0%; Pred. No. 2.9e-17;
Matches 127; Conservative 59; Mismatches 136; Indels 31; Gaps 10;

QY 19 AAVGLGATAPQAQAPALQSDRFAFPALPLPSA---MVAQVAPVNVNINIKLGYNNA 75
DB 102 AASLVGFNRPAGSGGPGVAASAPSIIPANMPGSGVEQVAAKVPVSVVWLETDLGRQSE 161
QY 76 VGAGTGVIDPENGVLNNHVIAGAT-----DINAFVSGGQTYGVVDVWGYDRTQDV 127
DB 162 --EGSGILSAEGLILNNHVIAAAKPPLGSPPKTTVTFSDGRTAPFTVVGADPTSDI 219
QY 128 AVQLRGAGGLPSAAIGG--VAVGEPVAMGNSGGGCTPRAVGRVVALGQTVQASDS 185
DB 220 AVRVQGVSGLTPIISLSSSDLRVQGVLAIGSPFLGLEGT--VTTGIVSALNRPVSTTGE 277
QY 186 LTGAETLNGLIQFDAAIQDGSQGVVNGLVGVGMNTA---ASDNFQLSQG--GQGF 239
DB 278 -AGNQNTVLDIAQDAINFNGSGGALVNNNAQLVGNVSAIATLGADSADAQSGSGLGFP 336
QY 240 AIPGQMAIAGQIRSGGSGFTVHIGPTAFGLGLGV--DNNNGARVQVGVGAPASLGI 298
DB 337 AIPVDQAKRIADELISTGKA-----SHASLGQVNTDKDTPGAKIVEVWAGGAANAGV 390
QY 299 STGDVITAVDGPAPNSATAMADALNGHHPGQDVISVNMQTKSGGTRTGNVTLAE 351
DB 391 PKGVVVTKVDDRPINSADALVAAVRSKAPGATVALTFQDPSSGGSRTVQVTLGK 443

RESULT 6
Q7U0X2 PRELIMINARY; PRT; 464 AA.
ID Q7U0X2
AC Q7U0X2;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DE Probable serine protease (Setine proteinase) (EC 3.4.21.-).
GN MB1009.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97; PubMed=12788972;
RX MEDLINE=22703107; PubMed=12788972;
RA Garnier T., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Pryor M., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA "The complete genome sequence of Mycobacterium bovis.";
RA Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
RL EMBL; BX248337; CAP93870.1;
RW Hydrolase; Complete proteome.
SQ SEQUENCE 464 AA; 46436 MW; AE9344B3FFA9B3 CRC64;

Query Match 26.1%; Score 469.5; DB 16; Length 446;
Best Local Similarity 36.0%; Pred. No. 2.9e-17;
Matches 127; Conservative 59; Mismatches 136; Indels 31; Gaps 10;

QY 19 AAVGLGATAPQAQAPALQSDRFAFPALPLPSA---MVAQVAPVNVNINIKLGYNNA 75
DB 120 AASLVGFNRPAGSGGPGVAASAPSIIPANMPGSGVEQVAAKVPVSVVWLETDLGRQSE 179

QY 76 VGAGTGVIDPENGVLNNHVIAGAT-----DINAFVSGGQTYGVVDVWGYDRTQDV 127
DB 180 --EGSGILSAEGLILNNHVIAAAKPPLGSPPKTTVTFSDGRTAPFTVVGADPTSDI 237
QY 128 AVQLRGAGGLPSAAIGG--VAVGEPVAMGNSGGGCTPRAVGRVVALGQTVQASDS 185
DB 238 AVRVQGVSGLTPIISLSSSDLRVQGVLAIGSPFLGLEGT--VTTGIVSALNRPVSTTGE 295
QY 186 LTGAETLNGLIQFDAAIQDGSQGVVNGLVGVGMNTA---ASDNFQLSQG--GQGF 239
DB 296 -AGNQNTVLDIAQDAINFNGSGGALVNNNAQLVGNVSAIATLGADSADAQSGSGLGFP 354
QY 240 AIPGQMAIAGQIRSGGSGFTVHIGPTAFGLGLGV--DNNNGARVQVGVGAPASLGI 298
DB 355 AIPVDQAKRIADELISTGKA-----SHASLGQVNTDKDTPGAKIVEVWAGGAANAGV 408
QY 299 STGDVITAVDGPAPNSATAMADALNGHHPGQDVISVNMQTKSGGTRTGNVTLAE 351
DB 409 PKGVVVTKVDDRPINSADALVAAVRSKAPGATVALTFQDPSSGGSRTVQVTLGK 461

RESULT 7
OS3896 PRELIMINARY; PRT; 464 AA.
ID OS3896;
AC OS3896;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Putative serine protease.
DE Putative serine protease.
GN RV0983 OR MTW044.11.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulten J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RA complete genome sequence.";
RA Nature 393:537-544(1998).
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
CC EMBL; AL021999; CAA17582.1;
DR PIR; C70821; C70821.
DR Tuberculin; RV0983;
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PS00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 464 AA; 46452 MW; AE93BFC53E1EC8F CRC64;

Query Match 26.0%; Score 468.5; DB 16; Length 464;
Best Local Similarity 36.0%; Pred. No. 3.3e-17;
Matches 127; Conservative 59; Mismatches 136; Indels 31; Gaps 10;

```


Query Match 25.5%; Score 460; DB 2; Length 452;
 Best Local Similarity 36.4%; Pred. No. 9e-17;
 Matches 130; Conservative 52; Mismatches 121; Indels 54; Gaps 12;

QY 22 GLGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQV-----POVNNINTKLGYNNAV 77
 DB 120 GAGPVTGPAASVPAANM-----PSGSVEQVAVKVPVSMLETLGRQSE-- 164

QY 78 ACTGIVDPNGVLTNNHVIAGA-----TDINAVSGGQTYGVVDVGVYDR 123
 DB 165 EGSVILSADGLILITNNHVAARAKPGCGGGGLSPKTTVTFF---DGRTPASTVVGADP 221

QY 124 TDQVAVLQLRGAGGLPSAAIIGGG--VAVGPPVVMGNSGGGTPRAVPGRVVVALGQTVQ 181
 DB 222 TSDIAVVRVQISGLTPTWSSADLRVGPVAVGSPPLAGLT--VTSGIVSALNRPV 279

QY 182 ASDSLTGAEETLNGLIQFDAAIQDGGSGPPVNGLGQVGMNTAA-----SDNFOLSQG 235
 DB 280 TTGSG-SGNQNTVLDATQDAINPNSGGALVNMGGQLVGNSAIAFLGADSGDAQSGSI 338

QY 236 GQGFAPIGQAMATAGQIRSGGSPVHIGPTAFLGLGVVDNNGN-CARQVRVVGSA 294
 DB 339 GLGFAIPVDQAKRIADLEITG--KATH-----ASLGQVATDGTGTPKARVMDVAGAAA 392

QY 295 SLGISTGDVITAVDGAPINSATAMADALNGHPDGVISVNMQTKSGGTRTGNVTLAE 351
 DB 393 NAAVPKGVLTQVDDRLISSADALVAARSKAPGDKVSLTYQDSGSSRTVQVTLCK 449

RESULT 10

Q9FBK9 PRELIMINARY; PRT; 542 AA.
 AC Q9FBK9;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative protease.
 GN SC05149 OR SCP8.12.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RA Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RA Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=37000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RA "A set of ordered cosmid and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,

RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RL coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 DR EMBL; AL939122; CAC01350.1;
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; P:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0008508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KW Hydrolase; Protease; Serine protease; Complete proteome.
 SQ SEQUENCE 542 AA; 53761 MW; 620F7D889DF7212D CRC64;

Query Match 22.6%; Score 407; DB 16; Length 542;
 Best Local Similarity 32.6%; Pred. No. 7e-14;
 Matches 116; Conservative 56; Mismatches 138; Indels 46; Gaps 12;

QY 22 GLGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVPOVNNINTKLGYNNAVAGAGTG 81
 DB 206 GVGTVELP-QAGPEAERD-----PDSVAGIAARALPSVTLH--VSGSEAAGTGTG 254

QY 82 IVIDPNGVLTNNHVI--AGATDINAFVSGGQTYGVVDVGVYDRDQVAVLQLRGAGCLP 139
 DB 255 FVLDRGHILTNHVEPAGSGEITVTNSGDTAEAEVWGRDSDGYLAVVKGVTGLT 314

QY 140 SAAITGG--GVAAGBPVVMGNSGGGTPRAVPGRVVVALGQTVQ--SDSLTGAEETLNG 196
 DB 315 PMLGNSDNVVRGDPVVAIGAPFLAGT--VTSGIIIAKERPIITAGGEGGSDISYVDA 372

QY 197 IQFDAAIQDGGSGPPVNGLGQVGMNTA-----ASDNFOLSGGQGFAPICQAM 247
 DB 373 LOTDAPINPGNSGGLLDARGAIGNSAIRSADSGSDGQAGSIGLGFALPINOQK 432

QY 248 AIAGQIRSGGSPVHIGPTAFLGLGV-VNHN--GNGARVQ-----RVVGSAPASL 296
 DB 433 RVABEL-----INTGKAHPVIGITLDMNYTGDGARVSAGKGGDGPVATTCGPGAKA 484

QY 297 GISTGDVITAVDGAPINSATAMADALNGHPDGVISVNMQTKSGGTRTGNVTLAE 352
 DB 485 GIKFGDVITAVDQGVHSGEELIVKTRHPRGDRLETLQRDGKETKVSILVLSGSG 540

RESULT 11
 Q93J30 PRELIMINARY; PRT; 519 AA.
 ID Q93J30
 AC Q93J30;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative protease.
 GN SC03977 OR SCBAC25E3.14.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Collins M., Harris D.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

[3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=37000351; PubMed=8843436;
RA Kadenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:777-96(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thompson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wierozorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939118; CAC44701.1; --
GO: GO:0008233; F:peptidase activity; IEA.
DR GO: GO:0004295; F:trypsin activity; IEA.
DR GO: GO:0007242; P:intracellular signaling cascade; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00555; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00020; Tryp_Spc; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 519 AA; 50327 MW; 5CB8D5F0CC19E428 CRC64;

Query Match 22.0%; Score 396; DB 16; Length 519;
Best Local Similarity 32.2%; Pred. No. 2.Se-13;
Matches 119; Conservative 54; Mismatches 148; Indels 48; Gaps 13;

Qy 14 LLSVLAAGVLG-----LATAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60
Db 159 LIALVAGGLGGGLGYTLARNDESGTTSASDTGGSVKEDAGTVA-----GVAAKAL 212

Qy 61 PQVYNINIKLYNNVAGAGTGIIDPNGVLTNNHVIAGATDINAFSV--GSGQTYGVVDV 118
Db 213 PSTVTIQAE--GSGEGGTGTGFVDFKEGHIVTNNHVVAAVDDGKLSATFPNGKDYDAEV 271

Qy 119 VGYDRTQDVAVLQRLGA--GGLPSAICGG--VAGFEPVAMGNSGGGGTTPRAVGRVVA 175
Db 272 VGHAAQGVDAVAVIKLENAPSDLKPLALGDSKVAVGDSITAIAGAPFGLSNT--VTGTLISA 329

Qy 176 LGQTVQASDSLTAETLNGLIQPDAAIQPDGSGPVVNGILGOVGMNTA----ASDNFQ 231
Db 330 KNPVASSDGSADSKASYMSALQTDASINPNSGGPLLDAGNVIGINSALQSTNGGFG 389

Qy 232 LSQCGQ---GPAIPGQMAIAGQIRSGGSGPTVHIQPTAFPLGLGVVDNNGNARV--QR 286
Db 390 TQAGSICLGLPAIPVNAKFAVQQLIKSGRPVAKIGASVSL-----EETNGAKLTEQG 444

Qy 287 VVGS-----APASLGIETGDTITAVDGPAPINATAMADALNGHPGDVISVNWQTKSG 340
Db 445 VGGSDPVEKGGPADAGLKPGDVTITKLDNRVIDSGPTLIGETWTKPGDEVTIYE--RGG 503

Qy 341 GTRIGNVTL 349
Db 504 KQHTAEVTL 512

RESULT 12
QSDG87 PRELIMINARY; PRT; 375 AA.
ID QSDG87;
AC QSDG87;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Serine protease.
GN TLR2436.
OS Synecococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
CP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriquchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005377; BAC09988.1; --
GO: GO:0008233; F:peptidase activity; IEA.
DR GO: GO:0004295; F:trypsin activity; IEA.
DR GO: GO:0007242; P:intracellular signaling cascade; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 375 AA; 39724 MW; 0D2602EB534915F5 CRC64;

Query Match 21.7%; Score 391; DB 16; Length 375;
Best Local Similarity 33.6%; Pred. No. 3.1e-13;
Matches 108; Conservative 49; Mismatches 114; Indels 50; Gaps 10;

Qy 35 PALSQDRFAD--FPALPLDPSAMVAQVAVFVNVNINIKLYNNVAGAGTGIIDPNGVLTN 93
Db 67 PULSDPFFRQPFGLALPQE-----DRLRGQSGFIIDPSGIVMTN 108

Qy 94 NHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQRLG--ACGLPSAICGG--VAVG 150
Db 109 AHVVSQADTVN--RLKDGKRVFEGVGVDEVSDLAIVKLKGVTEPLTAPLGDSEVRVG 167

Qy 151 EPVWAMGNSGGGGTTPRAVGRVVAIG--QTVQASDSLTAETLNGLIQPDAAIQGDS 208
Db 168 DNIAVAGPLGLDNT-----VTLGIISTLHRSQAQVIGIPDKLDFDTQDAINPGNS 219

Qy 209 GGPVNVGLGVVGMNTAASDNFQLSGGGGPAIPGQMAIAGQIRSGGSGTTHIG--P 266
Db 220 GGPLLNEAGEVIGINTA-----IRADAMGIGFAIPINKAKALQARLIRGEKIQAHTIGQM 275

Qy 267 TAFPLGLGVVDNNG-----NGARVQRVVGSAASLGIETGDTITAVDGPAPINSA 315
Db 276 TITTPAMAKENNANPNSPVILPEVNGVLVQLVFNTPPAKAGLRWGDVITAVDGPITSA 335

Qy 316 TAMADALNGHPGDVISVNWQ 336
Db 336 DQLQITIVDSAAVGVNLNLTQ 356

RESULT 13
P72780

```

ID P72780 PRELIMINARY; PRT; 394 AA.
AC P72780;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Protease HHOA.
GN HHOA OR SLL1679.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria, Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hikosawa M., Sugitara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shampo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; D90900; BA016795.1; -
DR MEROPS; S01.274; -
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR008353; Peptidase_S1B_tx.
DR InterPro; IPR008256; Peptidase_S1B_V8.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR01774; EXPOLFOXIN.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50106; PDZ; 1.
DR Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 394 AA; 41336 MW; AD59D94811B8F57B CRC64;

Query Match 21.2%; Score 382; DB 16; Length 394;
Best Local Similarity 31.6%; Pred. No. 9.9e-13;
Matches 116; Conservative 50; Mismatches 117; Indels 84; Gaps 14;

QY 13 WLLSVLAAGVLGLATAP-----QAAPPA-----LSQDRFADFPALPDPMVA 57
DB 13 YLAFAVGTAFGIANLPHAVAAADDLPAPVITAQASVPLTSESFV-----AAAVS 63

QY 58 QVAPQVNTK-----LGNNAV-----GAGTGIVIDPVG 90
DB 64 RSGPAVVIDETVTRTDPILDDFPQEFFRSFPVPRRIAGSGFIIDNSGII 123

QY 91 LTNNHVTAGTIDNFAFVSGQYGVVDVDTQDVAVLQIRGAG-GLPASAIG--GGV 147
DB 124 LTNAHVVDGASKV-VVTLRDGRTFDQVGRGTDEVTDLAVVQKIEPQGSALFPVAPLGTSSNL 182

QY 148 AVCEPVVAMNSGGQGTTPRVPGRVVALG--QTVQASDSLTAETLNGLTFQDAIOP 205
DB 183 QVGWIAVGNPVLNDT-----VTLGIILGSRSAQAQIPDKRVEFIQTDALNP 234

QY 206 GDSGGPVNGLGVGVGNTAASDNFQSGQGFALPIGOAMAIAGIIRSGGSPVTHIG 265
DB 235 GNSGGPILLNARGEVIGINTA---IRADATGIGFALPIDQAKAIONTLAAGGTVPHPYIG 290

QY 266 PTAFLGLGV---VDNNGN-----CARQVRVVGSAAPASLGISGVDTAVDGNP 311
DB 291 -VQMMNITVDQAQNNRNPSPPIIEVDGILVMRLPQTPAERAGIRRGDIVAVDGNP 349

```

```

QY 312 INSATAM 318
DB 350 ISDGARL 356

RESULT 14
ID Q82FM9 PRELIMINARY; PRT; 472 AA.
AC Q82FM9;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Putative protease.
GN SAV4223
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites."
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis."
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005038; BAC71935.1; -
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR Protease; Complete proteome.
SQ SEQUENCE 472 AA; 46086 MW; AE04FAA409A3B7E0 CRC64;

Query Match 21.2%; Score 382; DB 16; Length 472;
Best Local Similarity 32.2%; Pred. No. 1.2e-12;
Matches 102; Conservative 57; Mismatches 134; Indels 24; Gaps 10;

QY 53 SAMVAQVAPQVNTKGLYNNNAVAGTGIVDPNGVVLNNHVIAGATDINAFSV--GS 110
DB 157 AGVAAKALPSTVTIEAQSSSEG--GTGTGFVFDKQGHVITNNHVAEAVDGLTATFPD 215

QY 111 GQTYGVVDVYDRTDVAVLQLRGA--GLPSAAITGG--VAVGEPVAVMNSGGQGTTPR 167
DB 216 GKKNVAVVGHQAGYDVAVVVKLNKPNAPSDLOPLTLIGSDSKVAVGDSITAIAGPFLSNT-- 273

QY 168 AVPRGVVALGTVQASDSLTAETLNGLTFQDAIOPGDSGGPVNGLGVGVGNTA-- 225
DB 274 VTTGIIIAKRNFPVASSDSSSSKASYNALQTDASINFGSGGLPDLDAQSGVIGINSAIQ 333

QY 226 ASDNFQLSQGGQ-----GPAIPGQAMAIAGIIRSGGSPVTHIGTAFI-----CLGV 274
DB 334 SSSSGGLSGSGSGSIGLGFALPINQAKYVAQELIKTKGFPYVPVIGASVLSLEBTGGAKI 393

```

QY 275 VDNNGNGARVORVVGSGAPASLGISTGDTITAVDGPINSATAMADALNGHHPGDVISVN 334
Db 394 TEQGASGS--DAITPNGPAKAGLKGFDVITKLDDWVIDSGPTLGEIWTHTPGATVKLT 451
QY 335 WQTKSGGTRTGNVTLAE 351
Db 452 Y-TRDGKARTTDVLGE 457

RESULT 15
Q82IL8 PRELIMINARY; PRT; 619 AA.
AC Q82IL8
DT 01-JUN-2003 (TEMBLrel. 24, Created)
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Putative serine proteinase.
GN SAV3115.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis; deducing the ability of producing secondary
RT metabolites";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005033; BAC70826.1; -
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR003003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
KW Complete proteome.
SQ SEQUENCE 619 AA; 64251 MW; D2D3C72A67D3B3A6 CRC64;

Query Match 21.2%; Score 382; DB 16; Length 619;
Best Local Similarity 33.7%; Pred. No. 1.7e-12;
Matches 114; Conservative 53; Mismatches 127; Indels 44; Gaps 12;
QY 38 SQDRFADFPALPLDPSAMVAQVAFQVNVNINIKLGYNNNAVAGTGTIVDPNGVLTNNHVI 97
Db 298 SEERADSVA-----GIAARALPSVVTLVHK--GSAAGTGTGVLDGRGHILTNHHV 349
QY 98 --AGATDINAFSGGVGVVGVYDRTQDVAVLQLRGAGLPSAIGG--GVAVGEPV 153
Db 350 EPAGSGSEIVTFSGGETAKATVGRDSGYDLAVKVGSGGLKPMPLGNSDNVQVGDV 409
QY 154 VAMGNSGGGGTTPRAVGRVVALGQTVQA--SDSLTGAETLNGLIQFDAAIQPDSCGPV 212
Db 410 VAIGAPFDLANT--VTSGIISAKERPTAGGKGDGSDVSYVDALQTDAPINPNSGGL 467

QY 213 VNGLQVVGMTAA-----SDNFQLSQGG---GFAIPIGQAMAIAGQIRSGGSPVTHIG 265
Db 468 LDSKARVVGIINSAIRSDADSSDQSGQAGSLGLGFAIPVQAKRVABELINTGRATHPVIG 527
QY 266 PTAFLGLGVNDNNGNGARV-----QVVVGSAPAASLSTGDTITAVDGPINSATA 317
Db 528 VTLDN-----DYTGSGARVGTKSNDDGSPVTRGGGDRAGIQAGDVITEVDGQRIHSGEE 582
QY 318 MADALNGHHPGDVISVNWQTKSGGTRTGN---VTLAEG 352
Db 583 LIVKIRAHPRGDRALTVE-----RDGKEKPEVTLVLG 614

Search completed: June 22, 2004, 17:22:45
JOB time : 38.0268 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:44:25 ; Search time 40.5137 Seconds
(without alignments)
2301.458 Million cell updates/sec

Title: US-09-886-349A-4

Perfect score: 1698

Sequence: 1 MEHHHHAPPALSQDRFADF.....QTKSGGTRTGNVTLAEGPPA 330

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001s:*

5: Geneseq2002s:*

6: Geneseq2003as:*

7: Geneseq2003bs:*

8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1698	100.0	330	5 AAE29702	Aae29702 Mycobacte
2	1698	100.0	330	5 AAE17566	Aae17566 Mycobacte
3	1695	99.8	330	5 AAE29703	Aae29703 Mycobacte
4	1695	99.8	330	5 AAE17567	Aae17567 Mycobacte
5	1695	99.8	330	7 ADA26372	Ada26372 Mycobacte
6	1695	99.8	723	7 ADA26354	Ada26354 Mycobacte
7	1695	99.8	1010	7 ADA26356	Ada26356 Mycobacte
8	1691	99.6	330	7 ADA26371	Ada26371 Mycobacte
9	1645	96.9	355	2 AAY05000	Aay05000 Mycobacte
10	1645	96.9	355	4 AAG22137	Aag22137 Mycobacte
11	1645	96.9	355	4 AAG81110	Aag81110 Mycobacte
12	1645	96.9	379	2 AAY04830	Aay04830 Mycobacte
13	1634	96.2	355	2 AAW32367	Aaw32367 Mycobacte
14	1634	96.2	355	2 AAW32435	Aaw32435 Mycobacte
15	1634	96.2	355	2 AAW44307	Aaw44307 Mycobacte
16	1634	96.2	355	2 AAW1670	Aaw1670 M. tuberc
17	1634	96.2	355	2 AAY38972	Aay38972 M. tuberc
18	1634	96.2	355	2 AAY39109	Aay39109 M. tuberc
19	1634	96.2	355	4 AAU01890	Aau01890 M. tuberc
20	1634	96.2	355	5 AAE29701	Aae29701 Mycobacte
21	1634	96.2	355	5 AAE17565	Aae17565 Mycobacte
22	1003.5	59.1	1016	7 ADA26370	Ada26370 M. bovis
23	995	58.6	1154	7 ADA26368	Ada26368 Mycobacte
24	994.5	58.6	825	7 ADA26366	Ada26366 Mycobacte
25	993.5	58.5	813	7 ADA26367	Ada26367 Mycobacte

26	992	58.4	1022	7 ADA26369	Ada26369 Mycobacte
27	987.5	58.2	875	7 ADA26365	Ada26365 Mycobacte
28	987	58.1	596	2 AAY32070	Aay32070 Mycobacte
29	987	58.1	596	5 AAE29710	Aae29710 Mycobacte
30	987	58.1	596	5 AAE17574	Aae17574 Mycobacte
31	987	58.1	599	5 AAU74599	Aau74599 Antigenic
32	987	58.1	729	4 AAU22142	Aau22142 Ra12-H9-3
33	987	58.1	729	5 AAE29708	Aae29708 Mycobacte
34	987	58.1	729	5 AAE17572	Aae17572 Mycobacte
35	987	58.1	729	7 ADA26373	Ada26373 Mycobacte
36	987	58.1	930	5 AAE29731	Aae29731 Mycobacte
37	987	58.1	930	7 ADA26364	Ada26364 Mycobacte
38	984	58.0	729	5 AAE29709	Aae29709 Mycobacte
39	984	58.0	729	5 AAE17573	Aae17573 Mycobacte
40	984	58.0	729	7 ADA26374	Ada26374 Mycobacte
41	982	57.8	726	5 AAU74588	Aau74588 Antigenic
42	982	57.8	729	2 AAY32059	Aay32059 Mycobacte
43	981	57.8	195	5 AAE29704	Aae29704 Mycobacte
44	981	57.8	195	5 AAE17568	Aae17568 Mycobacte
45	676	39.8	231	2 AAY32071	Aay32071 Mycobacte

ALIGNMENTS

RESULT 1
AAE29702
ID AAE29702 standard; protein; 330 AA.
XX
AC AAE29702;
XX
DT 27-JAN-2003 (first entry)
XX
DE Mycobacterium tuberculosis mature Ra35 antigenic protein.
DE
KW Vaccine; immunity; diagnostic agent; gene therapy; Ra35 antigen.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT Misc-difference 182
FT /note= "Encoded by GAG"
PT Misc-difference 183
PT /note= "Encoded by GCG"
XX
PN WO200272792-A2.
XX
PD 19-SEP-2002.
XX
PF 13-MAR-2002; 2002WO-US008223.
XX
PR 13-MAR-2001; 2001US-0275837P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Brannon M, Guderian J;
XX
DR WPI: 2002-759844/82.
DR N-PSDB; AD47077.
XX
PT New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif, M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity against pathogenic microorganisms e.g. Leishmania and Mycobacterium tuberculosis.
XX
PS Disclosure; Page 79-80; 155pp; English.
XX
CC The invention relates to a recombinant nucleic acid molecule encoding a fusion polypeptide. The recombinant nucleic acid comprises a heterologous polynucleotide sequence encoding an antigen or an antigenic fragment from Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a polypeptide or its fragment. The Leishmania polynucleotide is selected from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention

CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is M.
 CC tuberculosis mature Ra35 (N-terminus of MTB32A; Ra35FL) antigenic protein
 XX
 SQ Sequence 330 AA;

Query Match 100.0%; Score 1698; DB 5; Length 330;
 Best Local Similarity 100.0%; Pred. No. 1.9e-122;
 Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSDQRFADFPALDPSAMVAQVGPQVNNINIKLGYNNVAGAGTGIVID 60
 Db 1 MHHHHHAPPALSDQRFADFPALDPSAMVAQVGPQVNNINIKLGYNNVAGAGTGIVID 60
 QY 61 PNGVLTNNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120
 Db 61 PNGVLTNNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120
 QY 121 GVAVGEPVVMNGSGGGGTTPRVPGRVVALGQTVQASDSLTAETLNGLIQFDDAIIQ 180
 Db 121 GVAVGEPVVMNGSGGGGTTPRVPGRVVALGQTVQASDSLTAETLNGLIQFDDAIIQ 180
 QY 181 GDSGGPVVNGLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 Db 181 GDSGGPVVNGLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 QY 241 PTAFLGLGVNDNNGNGARVQVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGH 300
 Db 241 PTAFLGLGVNDNNGNGARVQVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 2

AAE17566

ID AAE17566 standard; protein; 330 AA.

XX AAE17566;

AC AAE17566;

XX AAE17566;

DT 22-APR-2002 (first entry)

XX Mycobacterium sp. MTB32A (Ra35FL) mature protein.

XX Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB32A; Ra32FL protein.

XX Mycobacterium sp.

OS

FH Key Location/Qualifiers

FT Region 8..202

FT Misc-difference /note= "Ra35 N-terminal peptide"

FT FT

FT FT

FT FT

FT FT

FT FT

FT FT

FT FT

FT FT

FT FT

FT FT

FT FT

FT FT

FT FT

FT FT

FT FT

FT FT

FT FT

FT FT

FT FT

FT FT

PA (CORI-) CORIXA CORP.
 XX Skeiky Y, Reed S, Alderson M;
 XX WPI; 2002-147798/19.
 DR N-PSDB; AAD28336.
 XX
 PT Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 XX Claim 70; Fig 6; 136pp; English.

CC The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB32A (Ra32FL) mature protein
 XX
 SQ Sequence 330 AA;

Query Match 100.0%; Score 1698; DB 5; Length 330;

Best Local Similarity 100.0%; Pred. No. 1.9e-122;

Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSDQRFADFPALDPSAMVAQVGPQVNNINIKLGYNNVAGAGTGIVID 60

Db 1 MHHHHHAPPALSDQRFADFPALDPSAMVAQVGPQVNNINIKLGYNNVAGAGTGIVID 60

QY 61 PNGVLTNNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120

Db 61 PNGVLTNNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120

QY 121 GVAVGEPVVMNGSGGGGTTPRVPGRVVALGQTVQASDSLTAETLNGLIQFDDAIIQ 180

Db 121 GVAVGEPVVMNGSGGGGTTPRVPGRVVALGQTVQASDSLTAETLNGLIQFDDAIIQ 180

QY 181 GDSGGPVVNGLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240

Db 181 GDSGGPVVNGLGQVVGWNTAASDNFQLSQGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240

QY 241 PTAFLGLGVNDNNGNGARVQVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGH 300

Db 241 PTAFLGLGVNDNNGNGARVQVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGH 300

QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 3

AAE29703

ID AAE29703 standard; protein; 330 AA.

XX

XX AAE29703;

AC AAE29703;

XX 27-JAN-2003 (first entry)

XX

DE Mycobacterium tuberculosis Ra35FLMutSA mutant antigenic protein.
XX Vaccine; immunity; diagnostic agent; gene therapy; Ra35FLMutSA; mutant;
KW antigen; mutein.
XX Mycobacterium tuberculosis.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 183 /note= "Wild type Ser substituted with Ala"
FT FT
XX WO200272792-A2.
XX
XX 19-SEP-2002.
XX
XX 13-MAR-2002; 2002WO-US008223.
XX
XX 13-MAR-2001; 2001US-0275837P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Brannon M, Guderian J;
XX
XX WPI; 2002-759844/82.
XX N-PSDB; AAD47078.
XX
XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
PT tuberculosis.
XX
XX Disclosure; Page 81-82; 155pp; English.
XX
XX The invention relates to a recombinant nucleic acid molecule encoding a
CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
CC polynucleotide sequence encoding an antigen or an antigenic fragment from
CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
CC polypeptide or its fragment. The Leishmania polynucleotide is selected
CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
CC are used in methods for eliciting immune response in mammals. They are
CC useful as vaccines to elicit protective immunity against pathogenic
CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
CC polypeptides are used for enhancing the expression of polynucleotides, as
CC in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is M.
CC tuberculosis Ra35FLMutSA mutant antigenic protein
XX
XX Sequence 330 AA;

Query Match 99.8%; Score 1695; DB 5; Length 330;
Best Local Similarity 99.7%; Pred. No. 3.2e-122;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHAPPALQSDRPADFPALPLDPSAMVAQGVQVNVNTKLGNNVAGAGTGIVD 60
Db 1 MHHHHHAPPALQSDRPADFPALPLDPSAMVAQGVQVNVNTKLGNNVAGAGTGIVD 60
Qy 61 PNGVLTNNHVIAGATDINAFSGGQTYGVVYVGYDRTQDVAVLQLRGAGGLPSAAIGG 120
Db 61 PNGVLTNNHVIAGATDINAFSGGQTYGVVYVGYDRTQDVAVLQLRGAGGLPSAAIGG 120
Qy 121 GVAVGPFVWAMNGSGGGTTPRAVPGRVVVALGTQVQASDLTCAETLNLQFDAAIQP 180
Db 121 GVAVGPFVWAMNGSGGGTTPRAVPGRVVVALGTQVQASDLTCAETLNLQFDAAIQP 180
Qy 181 GDSGGPVNGLGVVGNNTAASDNFQLSQGGQGFAPIGQAMAIAGIRSGGSPVTHIG 240
Db 181 GDAGGPVNGLGVVGNNTAASDNFQLSQGGQGFAPIGQAMAIAGIRSGGSPVTHIG 240
Qy 241 PTAFLGLGVVNNNGNGARVQVVGSPAASLGISTGDIITAVDGPINSATAMADALNGH 300
Db 241 PTAFLGLGVVNNNGNGARVQVVGSPAASLGISTGDIITAVDGPINSATAMADALNGH 300

Qy 301 HPGDVISVTWTKSGGTGTGNVTLAEGPPA 330
Db 301 HPGDVISVTWTKSGGTGTGNVTLAEGPPA 330

RESULT 4

AAE17567
ID AAE17567 standard; protein; 330 AA.
XX
AC AAE17567;
XX
DT 22-APR-2002 (first entry)
XX
DE Mycobacterium species Ra35FL mature protein mutant (S183A), Ra35FLMutSA.
XX
XX Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; Ra35FLMutSA protein; mutant; mutein.
XX
XX Mycobacterium sp.
XX
XX Key Location/Qualifiers
FT Region 8.202
FT /note= "Ra35 N-terminal peptide"
FT Misc-difference 183
FT /note= "Wild type Ser substituted with Ala"
FT Region 199..330
FT /note= "Ra35 C-terminal peptide, Ra12"
XX
XX WO200198450-A2.
XX
XX 27-DEC-2001.
XX
XX 20-JUN-2001; 2001WO-US019959.
XX
XX 20-JUN-2000; 2000US-00597796.
PR 01-FEB-2001; 2001US-0285737P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Alderson M;
XX
XX WPI; 2002-147798/19.
DR N-PSDB; AAD28337.
XX
XX Composition comprising MTB39 antigen and MTB32A antigen from
PT Mycobacterium species, useful for eliciting immune response in a subject.
XX
XX Claim 73; Fig 6; 136pp; English.
XX
XX The present invention relates to fusion proteins containing at least two
CC Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected with
CC Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC Ra35FL mature protein mutant (S183A), Ra35FLMutSA
XX
XX Sequence 330 AA;

Query Match 99.8%; Score 1695; DB 5; Length 330;
 Best Local Similarity 99.7%; Pred. No. 3.2e-122;
 Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSDRFPALPLDPSAMVAQVGVVNTKLGYNNAVAGAGTGIVID 60
 DB 1 MHHHHHAPPALSDRFPALPLDPSAMVAQVGVVNTKLGYNNAVAGAGTGIVID 60

QY 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120
 DB 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120

QY 121 GVAIGEPPVAMGNSGGGQTPRAVPGRVVAGLQTVQASDSLTGAEETLNGLIQFDAAIQ 180
 DB 121 GVAIGEPPVAMGNSGGGQTPRAVPGRVVAGLQTVQASDSLTGAEETLNGLIQFDAAIQ 180

QY 181 GDSGGPVVNGLGQVVGMMNTAASDNFQLSQGGQGFPAIPIGQAMALAGQIRSGGSPVTHIG 240
 DB 181 GDAGGPVVGNGLGQVVGMMNTAASDNFQLSQGGQGFPAIPIGQAMALAGQIRSGGSPVTHIG 240

QY 241 PTAFLGLGVDDNNGNGARVQVVGSAAPASLGISTGDTGDTAVDGPINSATAMADALNGH 300
 DB 241 PTAFLGLGVDDNNGNGARVQVVGSAAPASLGISTGDTGDTAVDGPINSATAMADALNGH 300

QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 5
 ADA26372
 ID ADA26372 standard; protein; 330 AA.
 XX ADA26372;
 AC ADA26372;
 DT 20-NOV-2003 (first entry)
 DE Mycobacterium MTB32AmuSA protein.
 XX fusion polypeptide; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculostatic; gene therapy; vaccine; mutant; mutein.
 XX Synthetic.
 OS Mycobacterium sp.
 XX Key Location/Qualifiers
 FT Misc-difference 183
 FT /note= "Wild-type Ser substituted by Ala"
 XX WO2003070187-A2.
 XX 28-AUG-2003.
 XX 18-FEB-2003; 2003WO-US004903.
 XX 15-FEB-2002; 2002US-0357351P.
 XX (CORI-) CORIXA CORP.
 XX Skeiky Y, Guderian J, Reed S;
 XX WPI; 2003-697554/66.
 XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 PT and MTB39 antigens, with or without the MTB85A antigen, from a
 PT Mycobacterium species, useful for diagnosing, preventing and/or treating
 PT tuberculosis infection.
 XX Disclosure; Fig 19; 112pp; English.
 PS The invention relates to a novel nucleic acid encoding a fusion
 CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and

CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
 CC A polypeptide of the invention has tuberculostatic activity. A
 CC polynucleotide of the invention may have a use in gene therapy, and as a
 CC vaccine. The methods and compositions of the present invention are useful
 CC for diagnosing, preventing and/or treating tuberculosis infection. The
 CC present sequence is used in the exemplification of the invention.
 XX Sequence 330 AA;

Query Match 99.8%; Score 1695; DB 7; Length 330;
 Best Local Similarity 99.7%; Pred. No. 3.2e-122;
 Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSDRFPALPLDPSAMVAQVGVVNTKLGYNNAVAGAGTGIVID 60
 DB 1 MHHHHHAPPALSDRFPALPLDPSAMVAQVGVVNTKLGYNNAVAGAGTGIVID 60

QY 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120
 DB 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120

QY 121 GVAIGEPPVAMGNSGGGQTPRAVPGRVVAGLQTVQASDSLTGAEETLNGLIQFDAAIQ 180
 DB 121 GVAIGEPPVAMGNSGGGQTPRAVPGRVVAGLQTVQASDSLTGAEETLNGLIQFDAAIQ 180

QY 181 GDSGGPVVNGLGQVVGMMNTAASDNFQLSQGGQGFPAIPIGQAMALAGQIRSGGSPVTHIG 240
 DB 181 GDAGGPVVGNGLGQVVGMMNTAASDNFQLSQGGQGFPAIPIGQAMALAGQIRSGGSPVTHIG 240

QY 241 PTAFLGLGVDDNNGNGARVQVVGSAAPASLGISTGDTGDTAVDGPINSATAMADALNGH 300
 DB 241 PTAFLGLGVDDNNGNGARVQVVGSAAPASLGISTGDTGDTAVDGPINSATAMADALNGH 300

QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 6
 ADA26354
 ID ADA26354 standard; protein; 723 AA.
 XX ADA26354;
 AC ADA26354;
 DT 20-NOV-2003 (first entry)
 DE Mycobacterium MTB32-MTB39F fusion protein.
 XX fusion polypeptide; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculostatic; gene therapy; vaccine.
 XX Chimeric.
 OS Mycobacterium sp.
 XX WO2003070187-A2.
 XX 28-AUG-2003.
 XX 18-FEB-2003; 2003WO-US004903.
 XX 15-FEB-2002; 2002US-0357351P.
 XX (CORI-) CORIXA CORP.
 XX Skeiky Y, Guderian J, Reed S;
 XX WPI; 2003-697554/66.
 XX N-PSDB; ADA26353.
 XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 PT and MTB39 antigens, with or without the MTB85A antigen, from a
 PT Mycobacterium species, useful for diagnosing, preventing and/or treating
 PT tuberculosis infection.

XX	Claim 4; Fig 2; 112pp; English.
XX	The invention relates to a novel nucleic acid encoding a fusion
XX	polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC	MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC	A polypeptide of the invention has tuberculosstatic activity. A
CC	polynucleotide of the invention may have a use in gene therapy, and as a
CC	vaccine. The methods and compositions of the present invention are useful
CC	for diagnosing, preventing and/or treating tuberculosis infection. The
CC	present sequence is used in the exemplification of the invention.
XX	
SQ	Sequence 723 AA;
	Query Match 99.8%; Score 1695; DB 7; Length 723;
	Best Local Similarity 99.7%; Pred.No.8.6e-122;
	Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy	1 MHHHHHAPPALSDRFPADLPDPSAMVAQVGPQVWINTKLGYNNAVAGCTGIVID 60
Dd	1 MHHHHHAPPALSDRFPADLPDPSAMVAQVGPQVWINTKLGYNNAVAGCTGIVID 60
Qy	61 PNGVLTNHVIAGATDINAFSVSGGTGYVDVVGYDRTPDVAVLQRGAGGLPSAAIGG 120
Dd	61 PNGVLTNHVIAGATDINAFSVSGGTGYVDVVGYDRTPDVAVLQRGAGGLPSAAIGG 120
Qy	121 GVAVGEPPVAMNSGGGGTTPRAYPGRVVALGVTVQASDSLVTGAETTLNGLIQFDAAIQP 180
Dd	121 GVAVGEPPVAMNSGGGGTTPRAYPGRVVALGVTVQASDSLVTGAETTLNGLIQFDAAIQP 180
Qy	181 GDGGPPVYVNGLGQVVGMMNTAASNDFQLSQGGGFPAIPGQAWAIAGQIRSGGSPPTVHIG 240
Dd	181 GDAGPPVYVNGLGQVVGMMNTAASNDFQLSQGGGFPAIPGQAWAIAGQIRSGGSPPTVHIG 240
Qy	241 PTAFLGLGVVDNNGGARVQRVWGSAAPASLGISTGBVITAVDGPINSATAMADALNGH 300
Dd	241 PTAFLGLGVVDNNGGARVQRVWGSAAPASLGISTGBVITAVDGPINSATAMADALNGH 300
Qy	301 HPGDIVISVTWTKSGGTRTGNNVTLAEGPPA 330
Dd	301 HPGDIVISVTWTKSGGTRTGNNVTLAEGPPA 330
RESULT 7	
ADA26356	
ID	ADA26356 standard; protein; 1010 AA.
XX	ADA26356;
AC	
XX	
DT	20-NOV-2003 (first entry)
DE	Mycobacterium MTB-102F fusion protein.
XX	fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW	tuberculosis; tuberculostatic; gene therapy; vaccine.
OS	Chimeric.
OS	Mycobacterium sp.
XX	
PN	WO2003070187-A2.
XX	
PD	28-AUG-2003.
XX	
PF	18-FEB-2003; 2003WO-US004903.
XX	
PR	15-FEB-2002; 2002US-0357351P.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Skeiky Y, Guderman J, Reed S;
XX	
DR	WPI; 2003-697554/66.
DR	N-PSDB; ADA26355.

XX	New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT	and MTB39 antigens, with or without the MTB85A antigen, from a
PT	Mycobacterium species, useful for diagnosing, preventing and/or treating
PT	tuberculosis infection.
XX	
PS	Claim 6; Fig 4; 112pp; English.
XX	
CC	The invention relates to a novel nucleic acid encoding a fusion
CC	polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC	MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC	A polypeptide of the invention has tuberculostatic activity. A
CC	polynucleotide of the invention may have a use in gene therapy, and as a
CC	vaccine. The methods and compositions of the present invention are useful
CC	for diagnosing, preventing and/or treating tuberculosis infection. The
CC	present sequence is used in the exemplification of the invention.
XX	
SQ	Sequence 1010 AA;
	Query Match 99.8%; Score 1695; DB 7; Length 1010;
	Best Local Similarity 99.7%; Pred. No. 1.3e-121;
	Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy	1 MHHHHHHAPALSQDREADFPALPLDSPAMVAQVGPVVNNTKLGYNNAVAGTGIIVD 60
Dd	1 MHHHHHHAPALSQDRFADFALPLDSPAMVAQVGPVVNNTKLGYNNAVAGTGIIVD 60
Qy	61 PNGVLVTNNHVIAGATDINAFSVGSGGTGYVDVYVG YDRTQDVAVLQLRGAGLP SAAITGG 120
Dd	61 PNGVLVTNNHVIAGATDINAFSVGSGGTGYVDVYVG YDRTQDVAVLQLRGAGLP SAAITGG 120
Qy	121 GVAVGEPPVANGNSGGGGGTPRAVPGRVVALGQT VQAASDSLTGA EETLNGLI QFDAAIQP 180
Dd	121 GVAVGEPPVANGNSGGGGGTPRAVPGRVVALGQT VQAASDSLTGA EETLNGLI QFDAAIQP 180
Qy	181 GDSGGPPVYNGLGQVVGVMNTAASNDFOLSQGQGF PAIFIGQAMA IAGQIRSGGSP TTVHIG 240
Dd	181 GDAGGPPVYNGLGQVVGVMNTAASNDFOLSQGQGF PAIFIGQAMA IAGQIRSGGSP TTVHIG 240
Qy	241 PTAFLGIGVVDNNGNGARVORVVGSPAASLSIGSTG DVIITAVDCAPINSATAMADALNGH 300
Dd	241 PTAFLGIGVVDNNGNGARVORVVGSPAASLSIGSTG DVIITAVDCAPINSATAMADALNGH 300
Qy	301 HPGDVISVTWTKSGGTRTGNTVTLAEGPPA 330
Dd	301 HPGDVISVTWTKSGGTRTGNTVTLAEGPPA 330
RESULT 8	
ADA26371	
ID	ADA26371 standard; protein; 330 AA.
XX	
AC	ADA26371;
XX	
DT	20-NOV-2003 (first entry)
XX	
DE	Mycobacterium wild type MTB32A protein.
XX	
KW	fusion polypeptide; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
XW	tuberculosis; tuberculostatic; gene therapy; vaccine.

XX Skeiky Y, Guderian J, Reed S;
 XX WPI; 2003-697554/66.
 XX
 XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 PT and MTB39 antigens, with or without the MTB85A antigen, from a
 PT Mycobacterium species, useful for diagnosing, preventing and/or treating
 PT tuberculosis infection.
 XX
 PS Disclosure; Fig 19; 112pp; English.
 XX
 CC The invention relates to a novel nucleic acid encoding a fusion
 CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
 CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
 CC A polypeptide of the invention has tuberculostatic activity. A
 CC polynucleotide of the invention may have a use in gene therapy, and as a
 CC vaccine. The methods and compositions of the present invention are useful
 CC for diagnosing, preventing and/or treating tuberculosis infection. The
 CC present sequence is used in the exemplification of the invention.
 XX
 SQ Sequence 330 AA;
 Query Match 99.6%; Score 1691; DB 7; Length 330;
 Best Local Similarity 99.7%; Pred. No. 6.6e-122;
 Matches 329; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MHHEHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNVAGAGTGIVID 60
 DB 1 MHHEHHHAPPALSQDRFADFPALPLDPSAHVAQVGPQVNNINTKLGYNNVAGAGTGIVID 60
 QY 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGG 120
 DB 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGG 120
 QY 121 GVAVGEPVAMGNSGGGGTTPRAVPGRVVAGLQTVQASDSLTAETLGLIQTDAAIQ 180
 DB 121 GVAVGEPVAMGNSGGGGTTPRAVPGRVVAGLQTVQASDSLTAETLGLIQTDAAIQ 180
 QY 181 GDSCGPPVNLGQVVGNNMTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 DB 181 GDSCGPPVNLGQVVGNNMTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 QY 241 PTAFLGLGVVDNNGNGARVORVVGSAAPASLIGISTGDIVITAVDGPINSATAMADALNGH 300
 DB 241 PTAFLGLGVVDNNGNGARVORVVGSAAPASLIGISTGDIVITAVDGPINSATAMADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 RESULT 9
 AAY05000
 ID AAY05000 standard; protein; 355 AA.
 XX
 XX AAY05000;
 XX
 XX 06-JUL-1999 (first entry)
 XX
 XX Mycobacterium species protein sequence 50D.
 DE
 XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.
 KW
 XX Mycobacterium sp.
 OS
 XX WO9909186-A2.
 PN
 XX 25-FEB-1999.
 PD
 XX 14-AUG-1998; 98WO-FR001813.
 PP
 XX

PR 14-AUG-1997; 97FR-00010404.
 PR 11-SEP-1997; 97FR-00011325.
 XX
 XX (INSP) INST PASTEUR.
 XX
 XX Gicquel B, Portnoie D, Lim E, Pellicio V, Guigueno A;
 PI Goguet De La Salmoniere Y;
 PI
 XX WPI; 1999-181045/15.
 DR N-PSDB; AAX34251.
 DR
 XX
 XX Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in infection-associated
 PT protein expression.
 PT
 XX Claim 32; Fig 50D; 309pp; French.
 XX
 CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
 CC from various Mycobacterium species microorganisms. The encoding
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection
 CC
 XX
 SQ Sequence 355 AA;
 Query Match 96.9%; Score 1645; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 2.5e-118;
 Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNVAGAGTGIVDPNGVLT 67
 DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNVAGAGTGIVDPNGVLT 92
 QY 68 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEP 127
 DB 93 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEP 152
 QY 128 VVAMGNSGGGGTTPRAVPGRVVAGLQTVQASDSLTAETLGLIQTDAAIQPGDSGGPV 187
 DB 153 VVAMGNSGGGGTTPRAVPGRVVAGLQTVQASDSLTAETLGLIQTDAAIQPGDSGGPV 212
 QY 188 VNLGQVVGNNMTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGTAFGL 247
 DB 213 VNLGQVVGNNMTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGTAFGL 272
 QY 248 GVDNNGNGARVORVVGSAAPASLIGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS 307
 DB 273 GVDNNGNGARVORVVGSAAPASLIGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS 332
 QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
 DB 333 VTWQTKSGGTRTGNVTLAEGPPA 355
 RESULT 10
 AAO22137
 ID AAO22137 standard; protein; 355 AA.
 XX
 XX AAO22137;
 XX
 XX 03-OCT-2002 (first entry)
 XX
 XX Mycobacterium tuberculosis MTB32A protein.
 DE
 XX
 XX Ral2; serine protease antigen; MTB32A; Mycobacterium tuberculosis;
 KW vaccine; eukaryotic; prokaryotic origin; serum antibody; fusion protein;
 KW immunogen; cytokine.
 XX
 OS Mycobacterium tuberculosis.
 XX
 XX WO200125401-A2.
 PN
 XX

DT	04-SEP-2001	(first entry)	
DE	Mycobacterium tuberculosis	potential drug target protein SEQ ID 161.	
XX	Drug target; growth; organism viability; characterisation.		
XX	Mycobacterium tuberculosis.		
OS	Mycobacterium tuberculosis.		
XX	W0200135317-Al.		
XX	17-MAY-2001.		
PD	13-NOV-2000; 2000WO-US031152.		
PF	12-NOV-1999; 99US-0165086P.		
XX	12-NOV-1999; 99US-0165124P.		
PR	01-FEB-2000; 2000US-0175531P.		
XX	(REGC) UNIV CALIFORNIA.		
XX	Eisenberg D, Rotstein SH, Marcotte EM;		
PI	WPI; 2001-329193/34.		
XX	N-PSDB; AAH51961.		
DR	Identifying nucleotide or polypeptide sequence for use as drug target,		
DR	involves providing algorithm that analyzes a functional relationship		
PT	between nucleotide or polypeptide sequences, and comparing the sequences.		
PT	Disclosure; Page 157; 207pp; English.		
XX	This invention relates to a method for identifying a nucleotide or		
XX	polypeptide sequence that may be a drug target, or essential for growth		
CC	or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092		
CC	represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium		
CC	tuberculosis proteins which are potential drug targets. The DNA and		
CC	protein sequences are used to illustrate the method of the invention. The		
CC	method involves providing an unknown nucleotide or polypeptide sequences,		
CC	and comparing it to a number of sequences along with at least one		
CC	algorithm capable of analysing a functional relationship between		
CC	nucleotide and polypeptide sequences. The method is useful for		
CC	characterising the function of nucleic acids and polypeptides that may be		
CC	useful as a target for a drug or essential for the growth or viability of		
CC	an organism		
XX	Sequence 355 AA;		
XX	Query Match 96.9%; Score 1645; DB 4; Length 355;		
XX	Best Local Similarity 100.0%; Pred. No. 2.5e-118;		
XX	Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0		
Qy	8 APPALSQDRFADFPALFLDPSAMVAQVGPQVNNITKLGYNNAVGAGTGIVIDPNCVULT 67		
Db	33 APPALSQDRFADFPALFLDPSAMVAQVGPQVNNITKLGYNNAVGAGTGIVIDPNCVULT 92		
Qy	68 NNHVIAGATDINAFSVGSGQTYGVDDVGYDRTQDVAVLQLRGAGGLPSAAITGGGVAVGEP 127		
Db	93 NNHVIAGATDINAFSVGSGQTYGVDDVGYDRTQDVAVLQLRGAGGLPSAAITGGGVAVGEP 152		
Qy	128 VWAMNMSGGGGGTTPRAVGRVVALGQTVQASDSLTGABETTLNGLTFQDAAIQPGDSGGPV 187		
Db	153 VWAMNMSGGGGGTTPRAVGRVVALGQTVQASDSLTGABETTLNGLTFQDAAIQPGDSGGPV 212		
Qy	188 VNGLGQVVGWMTAASDNFOLSQCGGGAIPICQAWAIAQIRSGGSGTTHVIGPTAFJGL 247		
Db	213 VNGLGQVVGWMTAASDNFOLSQCGGGAIPICQAWAIAQIRSGGSGTTHVIGPTAFJGL 272		
Qy	248 GVDNNGNGARVQRVWGSAPASLIGISTGDVITAVDGAIPINSATAMADALNGHHPGDVIS 307		
Db	273 GVDNNGNGARVQRVWGSAPASLIGISTGDVITAVDGAIPINSATAMADALNGHHPGDVIS 332		
Qy	308 VTWQTKSGGTTGNVTIAEGPPA 330		

Db 333 VTWQKSGGTRTGNVTLAGPPA 355

RESULT 12

AAAY04830
ID AAY04830 standard; protein; 379 AA.

AC AAY04830;

DT 06-JUL-1999 (first entry)

XX Mycobacterium species protein sequence 50F.

XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW hybridisation; detection; vaccine; immunisation; infection.

XX Mycobacterium sp.

OS WO9909186-A2.

XX 25-FEB-1999.

XX 14-AUG-1998; 98WO-FR001813.

XX 14-AUG-1997; 97FR-00010404.

PR 11-SEP-1997; 97FR-00011325.

XX (INSP) INST PASTEUR.

XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
PI Goguet De La Salmoniere Y;

XX WPI; 1999-181045/15.

DR N-PSDB; AAX34252.

XX Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in infection-associated
PT protein expression.

XX Claim 32; Fig 50F; 309pp; French.

XX Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
CC from various Mycobacterium species microorganisms. The encoding
CC nucleotide sequences can be used as primers and probes for methods for
CC detecting and identifying mycobacteria, especially belonging to the M.
CC tuberculosis complex. The encoded proteins can be used in vaccines for
CC immunisation against a bacterial or viral infection

XX Sequence 379 AA;

Query Match 96.9%; Score 1645; DB 2; Length 379;

Best Local Similarity 100.0%; Pred. No. 2.7e-118;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPQVNTKLGYNNAVAGTGVIDPNCVLT 67

Db 57 APPALSQDRFADFPALPLDPSAMVAQVGPQVNTKLGYNNAVAGTGVIDPNCVLT 116

QY 68 NNHVIAGATDINAFSVGSGQTYGVVGYDRTQDVAVLQLRGAGGLPSAAGGVAVGEP 127

Db 117 NNHVIAGATDINAFSVGSGQTYGVVGYDRTQDVAVLQLRGAGGLPSAAGGVAVGEP 176

QY 128 VVAMGNSGGGGTTPRAVPGRVWALGQTVQASDSLTCGAETLNLGIQFDAAIQPGSGGPV 187

Db 177 VVAMGNSGGGGTTPRAVPGRVWALGQTVQASDSLTCGAETLNLGIQFDAAIQPGSGGPV 236

QY 188 VNLGLGVGMNTAASDNFQSLQGGQFAPIGQMAIAGQIRSGGSPVHIGPTAFGL 247

Db 237 VNLGLGVGMNTAASDNFQSLQGGQFAPIGQMAIAGQIRSGGSPVHIGPTAFGL 296

QY 248 GVVDDNNGNGARVQVRVGSAPAAASLGISTGVTAVDGPINSATAMADALNGHFGDVIS 307

Db 297 GVVDDNNGNGARVQVRVGSAPAAASLGISTGVTAVDGPINSATAMADALNGHFGDVIS 356

QY 308 VTWQKSGGTRTGNVTLAGPPA 330

Db 357 VTWQKSGGTRTGNVTLAGPPA 379

RESULT 13

AAW32367

ID AAW32367 standard; protein; 355 AA.

XX AAW32367;

XX 13-JAN-1998 (first entry)

DE Mycobacterium tuberculosis antigen TbRa35.

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.

XX Mycobacterium tuberculosis.

XX WO9709429-A2.

XX 13-MAR-1997.

XX 30-AUG-1996; 96WO-US014675.

XX 01-SEP-1995; 95US-00523435.

XX 22-SEP-1995; 95US-00532136.

XX 22-MAR-1996; 96US-00620280.

XX 05-JUN-1996; 96US-00658800.

XX 12-JUL-1996; 96US-006580573.

XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedwick TH, Twardzik DR;

XX WPI; 1997-192904/17.

DR N-PSDB; AAT91414.

XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
PT useful for diagnosis of M. tuberculosis infection.

PS Example 3; Page 124-126; 190pp; English.

XX A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
CC variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a M.tuberculosis antigen,
CC TbRa35. The immunogenic polypeptide can be used to diagnose
CC M.tuberculosis infection by forming complexes with specific antibodies in
CC the sample. Fragments of DNA encoding the immunogenic polypeptide can be
CC used as diagnostic primers or probes and agents that bind to the antigen,
CC especially monoclonal antibodies or equivalent polyclonal antibodies, are
CC also used for diagnosis

XX Sequence 355 AA;

Query Match 96.2%; Score 1634; DB 2; Length 355;

Best Local Similarity 99.4%; Pred. No. 1.7e-117;

Matches 32; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPQVNTKLGYNNAVAGTGVIDPNCVLT 67

Db 33 APPALSQDRFADFPALPLDPSAMVAQVGPQVNTKLGYNNAVAGTGVIDPNCVLT 92

QY 68 NNHVIAGATDINAFSVGSGQTYGVVGYDRTQDVAVLQLRGAGGLPSAAGGVAVGEP 127

Db 93 NNHVIAGATDINAFSVGSGQTYGVVGYDRTQDVAVLQLRGAGGLPSAAGGVAVGEP 152

QY 128 VVAMGNSGGGGTTPRAVPGRVWALGQTVQASDSLTCGAETLNLGIQFDAAIQPGSGGPV 187

Db 153 VVAMGNSGGGGTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
 QY 188 VNLGQVVGWMTAASDNFQSGGGGFAIPQAMAIAGQIRSGGSGPTVHIGPTAFGL 247
 Db 213 VNLGQVVGWMTAASDNFQSGGGGFAIPQAMAIAGQIRSGGSGPTVHIGPTAFGL 272
 QY 248 GVDNNGNGARVQRVVGSAASLGIISTGDTITAVDGPINSATAMADALNGHHPGDVIS 307
 Db 273 GVDNNGNGARVQRVVGSAASLGIISTGDTITAVDGPINSATAMADALNGHHPGDVIS 332
 QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
 Db 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 14
 AAW32435
 ID AAW32435 standard; protein; 355 AA.
 XX
 AC AAW32435;
 XX
 DT 08-JAN-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen TbrA35.
 XX
 KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 KW skin testing; M.tuberculosis.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9709428-A2.
 XX
 PD 13-MAR-1997.
 XX
 PF 30-AUG-1996; 96WO-US014674.
 XX
 PR 01-SEP-1995; 95US-00523436.
 PR 22-SEP-1995; 95US-00533634.
 PR 22-MAR-1996; 96US-00620874.
 PR 05-JUN-1996; 96US-00659683.
 PR 12-JUL-1996; 96US-00680574.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TH, Twardzik DR;
 XX
 DR WPI; 1997-192903/17.
 DR N-PSDB; AAT31477.
 XX
 XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also for
 PT diagnosis.
 XX
 PS Example 3; Page 114-116; 168pp; English.
 XX
 CC A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
 CC variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis antigen,
 CC TbrA35. The immunogenic protein, and fusion proteins containing one or
 CC more of the proteins or one of the proteins plus ESAT-6, are useful in
 CC vaccines, preferably when formulated with a non-specific adjuvant, to
 CC induce an immune response against M.tuberculosis (for treatment or
 CC prevention)
 XX
 SQ Sequence 355 AA;

Query Match 96.2%; Score 1634; DB 2; Length 355;
 Best Local Similarity 99.4%; Pred. No. 1.7e-117;
 Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 33 APPALSDQRFADFPALPLDPSAMVAQVAPQVWNTKLGYNNAVGAGTGIVDPNGVLT 92
 QY 68 NNHVIAGATDINAFSVSGGQTYGVVVGXYDRTODVAVLQLRGAGGLPSAAIGGSAVAYGEP 127
 Db 93 NNHVIAGATDINAFSVSGGQTYGVVVGXYDRTODVAVLQLRGAGGLPSAAIGGSAVAYGEP 152
 QY 128 VVAMGNSGGGGTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
 Db 153 VVAMGNSGGGGTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
 QY 188 VNLGQVVGWMTAASDNFQSGGGGFAIPQAMAIAGQIRSGGSGPTVHIGPTAFGL 247
 Db 213 VNLGQVVGWMTAASDNFQSGGGGFAIPQAMAIAGQIRSGGSGPTVHIGPTAFGL 272
 QY 248 GVDNNGNGARVQRVVGSAASLGIISTGDTITAVDGPINSATAMADALNGHHPGDVIS 307
 Db 273 GVDNNGNGARVQRVVGSAASLGIISTGDTITAVDGPINSATAMADALNGHHPGDVIS 332
 QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
 Db 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 15
 AAW64307
 ID AAW64307 standard; protein; 355 AA.
 XX
 AC AAW64307;
 XX
 DT 17-OCT-2003 (revised)
 DT 09-NOV-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen TbrA35.
 XX
 KW Tuberculosis; infection; diagnosis; antigen; TbrA35.
 XX
 OS Mycobacterium tuberculosis; strain H37Ra.
 XX
 PN WO9816645-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US018214.
 XX
 PR 11-OCT-1996; 96US-00729622.
 PR 13-MAR-1997; 97US-00818111.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX
 DR WPI; 1998-251292/22.
 DR N-PSDB; AAV44355.
 XX
 PT New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and
 PT diagnosis of tuberculosis.
 XX
 PS Example 3; Page 115-116; 250pp; English.
 XX
 CC This polypeptide comprises Mycobacterium tuberculosis soluble antigen
 CC TbrA35. It is encoded by a DNA sequence (see AAV44355) isolated from a M.
 CC tuberculosis strain H37Ra expression library with rabbit anti-sera raised
 CC against M. tuberculosis supernatant. No significant homology was found
 CC between TbrA35 and Genbank database sequences. The invention relates to
 CC compositions and methods for diagnosing tuberculosis. It provides
 CC polypeptides (see AAW64291-W64379) comprising an antigenic portion of a
 CC soluble M. tuberculosis antigen, or an immunogenic portion of an M.
 CC tuberculosis antigen, as well as DNA sequences encoding such
 CC polypeptides, recombinant expression vectors and transformed or
 CC transfected host cells. Also claimed are methods and diagnostic kits for

CC detecting M. tuberculosis infection in a patient using the above
CC polypeptides, antibodies or oligonucleotide probes and primers, for the
CC diagnosis of tuberculosis. (Updated on 17-Oct-2003 to standardise OS
CC field)

XX
SQ Sequence 355 AA;

Query Match	96.2%;	Score 1634;	DB 2;	Length 355;
Best Local Similarity	99.4%;	Pred. No. 1.7e-117;	Indels 0;	Gaps 0;
Matches 321;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;

QY	8	APPALSDQRFADFPALPDPFSAMVAQVPQVNNINTKLGYNNAVAGTGIVIDPNGVLT	67
Db	33	APPALSDQRFADFPALPDPFSAMVAQVPQVNNINTKLGYNNAVAGTGIVIDPNGVLT	92
QY	68	NNHVIAGATDINAFSVSGGQTYGVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP	127
Db	93	NNHVIAGATDINAFSVSGGQTYGVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP	152
QY	128	VVAMNSGGGGTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQFGDSGGPV	187
Db	153	VVAMNSGGGGTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQFGDSGGPV	212
QY	188	VNGLGQVVGWNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIGTAFGL	247
Db	213	VNGLGQVVGWNTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIGTAFGL	272
QY	248	GVVDNNGNGARVQRVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS	307
Db	273	GVVDNNGNGARVQRVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS	332
QY	308	VTWQTKSGGTRTGNVTLAEGPPA	330
Db	333	VNWQTKSGGTRTGNVTLAEGPPA	355

Search completed: June 22, 2004, 17:15:24
Job time : 41.5137 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:07:54 ; Search time 11.78 Seconds
(without alignments)
1446.225 Million cell updates/sec

Title: US-09-886-349A-4
Perfect score: 1698
Sequence: 1 MSHHHHAPPALSQDRFADF.....QTKSGGTRGNVTLAGGPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/prodata/2/iaa/PCUTUS_COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1634	96.2	355	3	US-08-818-112-79
2	1634	96.2	355	4	US-08-818-111-80
3	1634	96.2	355	4	US-09-056-556-79
4	1634	96.2	355	4	US-09-072-596-80
5	1634	96.2	355	4	US-09-072-967-79
6	987	58.1	596	4	US-09-287-849-26
7	987	58.1	729	4	US-09-223-040-2
8	987	58.1	729	4	US-09-287-849-2
9	676	39.8	231	4	US-09-287-849-28
10	670	39.5	132	3	US-08-818-112-66
11	670	39.5	132	4	US-08-818-111-67
12	670	39.5	132	4	US-09-056-556-66
13	670	39.5	132	4	US-09-072-596-67
14	670	39.5	132	4	US-09-072-967-66
15	670	39.5	132	4	US-09-836-215-819
16	670	39.5	132	4	US-09-836-215-848
17	670	39.5	132	4	US-09-885-166A-819
18	670	39.5	132	4	US-09-885-166A-848
19	633	37.3	224	4	US-09-836-215-825
20	633	37.3	224	4	US-09-885-166A-825
21	633	37.3	267	4	US-09-843-597-352
22	633	37.3	267	4	US-09-806-421B-352
23	633	37.3	273	4	US-09-736-457-1864
24	633	37.3	229	4	US-09-643-597-354
25	633	37.3	229	4	US-09-606-421B-354
26	633	37.3	304	4	US-09-636-215-835
27	633	37.3	304	4	US-09-685-166A-835

US-08-818-112-79

Sequence 79, Application US/08818112

Patent No. 6290969

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.

APPLICANT: Campos-Neto, Antonio

APPLICANT: Houghton, Raymond

APPLICANT: Vedvick, Thomas S.

APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

NUMBER OF SEQUENCES: 153

TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/818.112

FILING DATE: 13-MAR-1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.411C6

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 79:

SEQUENCE CHARACTERISTICS:

LENGTH: 355 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-818-112-79

Query Match 96.2%; Score 1634; DB 3; Length 355;

Best Local Similarity 99.4%; Pred. No. 2.6e-131;

Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Sequence 1863, App

Sequence 852, App

Sequence 852, App

Sequence 349, App

Sequence 349, App

Sequence 333, App

Sequence 333, App

Sequence 196, App

Sequence 196, App

Sequence 196, App

Sequence 353, App

Sequence 353, App

Sequence 337, App

Sequence 337, App

Sequence 309, App

Sequence 309, App

Sequence 325, App

Sequence 325, App

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVNNINTKLGYNNAVAGAGTGVIVDPNGVILT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVNNINTKLGYNNAVAGAGTGVIVDPNGVILT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVVGYDRTPQVAVLQIRGAGGLPSAAIGGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVVGYDRTPQVAVLQIRGAGGLPSAAIGGGVAVGEP 152
QY 128 VVAMNGSGGGTTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDDAIQPGDSGGPV 187
DB 153 VVAMNGSGGGTTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDDAIQPGDSGGPV 212
QY 188 VNLGQVVGMMNTAASDNFQLSQGGQGFPAIPGQAMAIAGQIRSGGSPVTHIGPTAFGL 247
DB 213 VNLGQVVGMMNTAASDNFQLSQGGQGFPAIPGQAMAIAGQIRSGGSPVTHIGPTAFGL 272
QY 248 GVVDNNGGARVQVVGSAASIGISTGDDVITAVDGAPINSATAMADALNGHHHPGDVIS 307
DB 273 GVVDNNGGARVQVVGSAASIGISTGDDVITAVDGAPINSATAMADALNGHHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VNMQTKSGGTRTGNVTLAEGPPA 355

RESULT 2

US-08-818-111-80
; Sequence 80, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-80

Query Match 96.2%; Score 1634; DB 4; Length 355;
Best Local Similarity 99.4%; Pred. No. 2.6e-131;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVNNINTKLGYNNAVAGAGTGVIVDPNGVILT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVNNINTKLGYNNAVAGAGTGVIVDPNGVILT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVVGYDRTPQVAVLQIRGAGGLPSAAIGGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVVGYDRTPQVAVLQIRGAGGLPSAAIGGGVAVGEP 152
QY 128 VVAMNGSGGGTTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDDAIQPGDSGGPV 187
DB 153 VVAMNGSGGGTTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDDAIQPGDSGGPV 212
QY 188 VNLGQVVGMMNTAASDNFQLSQGGQGFPAIPGQAMAIAGQIRSGGSPVTHIGPTAFGL 247
DB 213 VNLGQVVGMMNTAASDNFQLSQGGQGFPAIPGQAMAIAGQIRSGGSPVTHIGPTAFGL 272
QY 248 GVVDNNGGARVQVVGSAASIGISTGDDVITAVDGAPINSATAMADALNGHHHPGDVIS 307
DB 273 GVVDNNGGARVQVVGSAASIGISTGDDVITAVDGAPINSATAMADALNGHHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VNMQTKSGGTRTGNVTLAEGPPA 355

RESULT 3

US-09-056-556-79
; Sequence 79, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-79

Query Match 96.2%; Score 1634; DB 4; Length 355;
Best Local Similarity 99.4%; Pred. No. 2.6e-131;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVNNINTKLGYNNAVAGAGTGVIVDPNGVILT 67

Db 33 APPALSQDRFADFPALPLDPSAMVAQVAPQVNNINTKLGYNNAVAGAGTGIVIDPNGVULT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVYVGYDRQDVAVLQLRGAGGLPSAAIGGVAVGEP 127
Db 93 NNHVIAGATDINAFSVSGGQTYGVYVGYDRQDVAVLQLRGAGGLPSAAIGGVAVGEP 152
QY 128 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
Db 153 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
QY 188 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 247
Db 213 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 272
QY 248 GVVDNNGNGARVQVVGSAAPASLGI STGDVITAVDGPINSATAMADALNGHHFGDVIS 307
Db 273 GVVDNNGNGARVQVVGSAAPASLGI STGDVITAVDGPINSATAMADALNGHHFGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
Db 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 4
US-09-072-596-80
; Sequence 80, Application US/09072596
; Patent No. 6458365
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-80

Query Match 96.2%; Score 1634; DB 4; Length 355;
Best Local Similarity 99.4%; Pred. No. 2.6e-131;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVAPQVNNINTKLGYNNAVAGAGTGIVIDPNGVULT 67
Db 33 APPALSQDRFADFPALPLDPSAMVAQVAPQVNNINTKLGYNNAVAGAGTGIVIDPNGVULT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVYVGYDRQDVAVLQLRGAGGLPSAAIGGVAVGEP 127
Db 93 NNHVIAGATDINAFSVSGGQTYGVYVGYDRQDVAVLQLRGAGGLPSAAIGGVAVGEP 152
QY 128 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
Db 153 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
QY 188 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 247
Db 213 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 272
QY 248 GVVDNNGNGARVQVVGSAAPASLGI STGDVITAVDGPINSATAMADALNGHHFGDVIS 307
Db 273 GVVDNNGNGARVQVVGSAAPASLGI STGDVITAVDGPINSATAMADALNGHHFGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
Db 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 5
US-09-072-967-79
; Sequence 79, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-967-79

Query Match 96.2%; Score 1634; DB 4; Length 355;
Best Local Similarity 99.4%; Pred. No. 2.6e-131;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 8 APPALSQDRFADFPALPDSAMVAQVGPVNNVNTKLGYNNAVGAGTGIVDPNGVWLT 67
Db 33 APPALSQDRFADFPALPDSAMVAQVGPVNNVNTKLGYNNAVGAGTGIVDPNGVWLT 92
Qy 68 NNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 127
Db 93 NNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 152
Qy 128 VVAMNSGGGGTTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 187
Db 153 VVAMNSGGGGTTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 212
Qy 188 VNLGQVVGMMNTAASDFQISQGGQFAIPIGQAMATAGQIRSGGSPVTHIGTFAFLGL 247
Db 213 VNLGQVVGMMNTAASDFQISQGGQFAIPIGQAMATAGQIRSGGSPVTHIGTFAFLGL 272
Qy 248 GVDNNGNGARVQVVGSAAPASIGISTGDIITAVDGAIPINSATAMADALNGHHGPDVIS 307
Db 273 GVDNNGNGARVQVVGSAAPASIGISTGDIITAVDGAIPINSATAMADALNGHHGPDVIS 332
Qy 308 VTWQTKSGGTRTGNVTLAEGPPA 330
Db 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 6

US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. 6627198

GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Alderson, Mark

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Fusion proteins of Mycobacterium tuberculosis Antigens

; FILE REFERENCE: 014058-009020US

; CURRENT FILING DATE: 1999-04-07

; PRIOR FILING DATE: 1997-03-13

; PRIOR FILING DATE: 1997-03-13

; PRIOR FILING DATE: 1997-10-01

; PRIOR FILING DATE: 1998-02-18

; PRIOR FILING DATE: 1998-02-18

; PRIOR FILING DATE: 1998-04-07

; PRIOR FILING DATE: 1998-04-07

; NUMBER OF SEQ ID NOS: 46

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 26

; LENGTH: 536

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: bi-fusion

US-09-287-849-26

Query Match 58.1%; Score 987; DB 4; Length 596;
Best Local Similarity 100.0%; Pred. No. 6.1e-76;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 APPALSQDRFADFPALPDSAMVAQVGPVNNVNTKLGYNNAVGAGTGIVDPNGVWLT 67
Db 402 APPALSQDRFADFPALPDSAMVAQVGPVNNVNTKLGYNNAVGAGTGIVDPNGVWLT 461

Qy 68 NNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 127
Db 462 NNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 521
Qy 128 VVAMNSGGGGTTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 187
Db 522 VVAMNSGGGGTTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 581
Qy 188 VNLGQVVGMMNTAAS 202
Db 582 VNLGQVVGMMNTAAS 596

RESULT 7

US-09-223-040-2

; Sequence 2, Application US/09223040

; Patent No. 654322

GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir

; APPLICANT: Alderson, Mark

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens

; TITLE OF INVENTION: and Their Uses

; FILE REFERENCE: 014058-009010US

; CURRENT APPLICATION NUMBER: US/09/223.040

; CURRENT FILING DATE: 1998-12-30

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 2

; LENGTH: 729

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion

US-09-223-040-2

Query Match 58.1%; Score 987; DB 4; Length 729;

Best Local Similarity 100.0%; Pred. No. 7.9e-76;

Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 APPALSQDRFADFPALPDSAMVAQVGPVNNVNTKLGYNNAVGAGTGIVDPNGVWLT 67
Db 535 APPALSQDRFADFPALPDSAMVAQVGPVNNVNTKLGYNNAVGAGTGIVDPNGVWLT 594

Qy 68 NNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 127
Db 595 NNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 654

Qy 128 VVAMNSGGGGTTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 187
Db 655 VVAMNSGGGGTTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 714

Qy 188 VNLGQVVGMMNTAAS 202

Db 715 VNLGQVVGMMNTAAS 729

RESULT 8

US-09-287-849-2

; Sequence 2, Application US/09287849

; Patent No. 6627198

GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Alderson, Mark

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens

; TITLE OF INVENTION: and Their Uses

; FILE REFERENCE: 014058-009020US

```
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-2

Query Match          58.1%; Score 987; DB 4; Length 729;
Best Local Similarity 100.0%; Pred. No. 7.9e-76;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 APPALSQDRPADPALPLDPSAMVAQVGPVNVNITKLYNNNAVAGAGTGIVIDPNCVILT 67
DB      535 APPALSQDRPADPALPLDPSAMVAQVGPVNVNITKLYNNNAVAGAGTGIVIDPNCVILT 594

QY      68 NNHVIAGATDINAFSVGSGGTGYGVVDVDRDQVAVLQLRGAGGLPSAAIGGVAVGEP 127
DB      595 NNHVIAGATDINAFSVGSGGTGYGVVDVDRDQVAVLQLRGAGGLPSAAIGGVAVGEP 654

QY      128 VVAMGNSGGCGTPRAVPGRVVAVLQGVQVSDSLTCAETLNGLIQFDDAIQPGDSGGPV 187
DB      655 VVAMGNSGGCGTPRAVPGRVVAVLQGVQVSDSLTCAETLNGLIQFDDAIQPGDSGGPV 714

QY      188 VVGLGQVVGMMNTAAS 202
DB      715 VVGLGQVVGMMNTAAS 729

RESULT 9
US-09-287-849-28
; Sequence 28, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-28

Query Match          39.8%; Score 676; DB 4; Length 231;
Best Local Similarity 99.2%; Pred. No. 5.2e-50;
Matches 132; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      198 NTAASDNFQLSQGGGQFAPIGQAVAIAGQIRSGGSGPTVHIGTAFGLGVVDNNGGA 257
DB      8 HTAASDNFQLSQGGGQFAPIGQAVAIAGQIRSGGSGPTVHIGTAFGLGVVDNNGGA 67

QY      258 RVQRVVGSAAPASLIGISTGDIVITAVDGPINSATAMADALNGHHHPGDIVISVTWTKSGGT 317
DB      68 RVQRVVGSAAPASLIGISTGDIVITAVDGPINSATAMADALNGHHHPGDIVISVTWTKSGGT 127

QY      318 RTGNVTLAEGPPA 330
DB      128 RTGNVTLAEGPPA 140

RESULT 10
US-08-818-112-66
; Sequence 66, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-66

Query Match          39.5%; Score 670; DB 3; Length 132;
Best Local Similarity 99.2%; Pred. No. 8e-50;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 199 TAASDNFQLSQGGGFAIPGQAWAIAQIRSGGSPVTHIGPTAFGLGVDNNGNGAR 258
DB 1 TAASDNFQLSQGGGFAIPGQAWAIAQIRSGGSPVTHIGPTAFGLGVDNNGNGAR 60
QY 259 VORVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNGHPGDVISTVWTKSGGTR 318
DB 61 VORVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNGHPGDVISTVWTKSGGTR 120
QY 319 TGNVTLAEGPPA 330
DB 121 TGNVTLAEGPPA 132

RESULT 11
US-08-818-111-67
; Sequence 67, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-67

Query Match 39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 8e-50;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 199 TAASDNFQLSQGGGFAIPGQAWAIAQIRSGGSPVTHIGPTAFGLGVDNNGNGAR 258
DB 1 TAASDNFQLSQGGGFAIPGQAWAIAQIRSGGSPVTHIGPTAFGLGVDNNGNGAR 60
QY 259 VORVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNGHPGDVISTVWTKSGGTR 318
DB 61 VORVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNGHPGDVISTVWTKSGGTR 120
QY 319 TGNVTLAEGPPA 330
DB 121 TGNVTLAEGPPA 132

RESULT 12
US-09-056-556-66
; Sequence 66, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056.556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-66

Query Match 39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 8e-50;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 199 TAASDNFQLSQGGGFAIPGQAWAIAQIRSGGSPVTHIGPTAFGLGVDNNGNGAR 258
DB 1 TAASDNFQLSQGGGFAIPGQAWAIAQIRSGGSPVTHIGPTAFGLGVDNNGNGAR 60
QY 259 VORVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNGHPGDVISTVWTKSGGTR 318
DB 61 VORVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNGHPGDVISTVWTKSGGTR 120
QY 319 TGNVTLAEGPPA 330
DB 121 TGNVTLAEGPPA 132

RESULT 13
US-09-072-596-67
; Sequence 67, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.

APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 622-6031
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-596-66

Query Match 39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 8e-50;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 199 TAASDNFQLSQGGQFAIPGQMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGGAR 258
Db 1 TAASDNFQLSQGGQFAIPGQMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGGAR 60

QY 259 VQRVVGSAAPASLGISTGVDITAVDGPINSATAMADALNGHHPGDVISVTWQTKSGGTR 318
Db 61 VQRVVGSAAPASLGISTGVDITAVDGPINSATAMADALNGHHPGDVISVTWQTKSGGTR 120

QY 319 TGNVTLAEGPPA 330
Db 121 TGNVTLAEGPPA 132

RESULT 14
US-09-072-967-66
Sequence 66, Application US/09072967
Patent No. 6592877
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington

COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 622-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-967-66

Query Match 39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 8e-50;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 199 TAASDNFQLSQGGQFAIPGQMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGGAR 258
Db 1 TAASDNFQLSQGGQFAIPGQMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGGAR 60

QY 259 VQRVVGSAAPASLGISTGVDITAVDGPINSATAMADALNGHHPGDVISVTWQTKSGGTR 318
Db 61 VQRVVGSAAPASLGISTGVDITAVDGPINSATAMADALNGHHPGDVISVTWQTKSGGTR 120

QY 319 TGNVTLAEGPPA 330
Db 121 TGNVTLAEGPPA 132

RESULT 15
US-09-636-215-819
Sequence 819, Application US/09636215
Patent No. 6620922
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqui
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.42717C17
CURRENT APPLICATION NUMBER: US/09/636,215
CURRENT FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 852
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 819

```
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-636-215-819

Query Match      39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 8e-50;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 199 TAASDNFOLSQGGGFAIPICQAWAIAGQIRSGGSPVTHIGPTAFILGLGVVDNNGAR 258
      |||||U|||||U|||||U|||||U|||||U|||||U|||||U|||||U|||||
Db 1 TAASDNFOLSQGGGFAIPICQAWAIAGQIRSGGSPVTHIGPTAFILGLGVVDNNGAR 60
      |||||U|||||U|||||U|||||U|||||U|||||U|||||U|||||

Qy 259 VQVVGSAPAASLGIISTGDVITAVDGPINSATAMADALNGHFGDVISVTWQTKSGGTR 318
      |||||U|||||U|||||U|||||U|||||U|||||U|||||U|||||
Db 61 VQVVGSAPAASLGIISTGDVITAVDGPINSATAMADALNGHFGDVISVNWQTKSGGTR 120
      |||||U|||||U|||||U|||||U|||||U|||||U|||||U|||||

Qy 319 TCNVTLAEGPPA 330
      |||||U|||||
Db 121 TCNVTLAEGPPA 132
      |||||U|||||
```

Search completed: June 22, 2004, 17:27:28
Job time : 11.78 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 30.5644 Seconds

(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-4
Perfect score: 1698
Sequence: 1 MHHHHHAPPALSQDRFADF.....QTKSGGTRGNVTLAEGPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1698	100.0	330	12	US-09-886-349A-4
2	1698	100.0	330	14	US-10-098-732A-4
3	1698	100.0	330	15	US-10-369-983-19
4	1695	99.8	330	12	US-09-886-349A-6
5	1695	99.8	330	14	US-10-098-732A-6
6	1695	99.8	330	15	US-10-369-983-20
7	1695	99.8	723	15	US-10-369-983-2
8	1695	99.8	1010	15	US-10-369-983-4
9	1645	96.9	355	9	US-09-712-363-161
10	1634	96.2	355	12	US-09-886-349A-2
11	1634	96.2	355	14	US-10-193-002-80
12	1634	96.2	355	14	US-10-084-843-79
13	1634	96.2	355	14	US-10-098-732A-2
14	1003.5	59.1	1016	15	US-10-369-983-18
15	995	58.6	1154	15	US-10-369-983-16

16	994.5	58.6	825	15	US-10-369-983-14	Sequence 14, Appl
17	993.5	58.5	813	15	US-10-369-983-15	Sequence 15, Appl
18	992	58.4	1022	15	US-10-369-983-17	Sequence 17, Appl
19	987.5	58.2	875	15	US-10-369-983-13	Sequence 13, Appl
20	987	58.1	596	9	US-09-287-849-26	Sequence 26, Appl
21	987	58.1	596	12	US-09-886-349A-20	Sequence 20, Appl
22	987	58.1	596	14	US-10-359-460-26	Sequence 26, Appl
23	987	58.1	596	14	US-10-098-732A-20	Sequence 20, Appl
24	987	58.1	729	9	US-09-287-849-2	Sequence 2, Appl
25	987	58.1	729	12	US-09-886-349A-16	Sequence 16, Appl
26	987	58.1	729	14	US-10-359-460-2	Sequence 2, Appl
27	987	58.1	729	14	US-10-098-732A-16	Sequence 16, Appl
28	987	58.1	729	15	US-10-369-983-21	Sequence 21, Appl
29	987	58.1	729	15	US-10-359-459-2	Sequence 2, Appl
30	987	58.1	930	14	US-10-098-732A-65	Sequence 65, Appl
31	987	58.1	930	15	US-10-369-983-12	Sequence 12, Appl
32	984	58.0	729	12	US-09-886-349A-18	Sequence 18, Appl
33	984	58.0	729	14	US-10-098-732A-18	Sequence 18, Appl
34	984	58.0	729	15	US-10-369-983-22	Sequence 22, Appl
35	981	57.8	195	12	US-09-886-349A-8	Sequence 8, Appl
36	981	57.8	195	14	US-10-098-732A-8	Sequence 8, Appl
37	676	39.8	231	9	US-09-287-849-28	Sequence 28, Appl
38	676	39.8	231	14	US-10-359-460-28	Sequence 819, App
39	670	39.5	132	9	US-09-759-143-819	Sequence 819, App
40	670	39.5	132	9	US-09-759-143-848	Sequence 819, App
41	670	39.5	132	9	US-09-780-669-819	Sequence 819, App
42	670	39.5	132	9	US-09-780-669-848	Sequence 819, App
43	670	39.5	132	9	US-09-822-827-819	Sequence 819, App
44	670	39.5	132	9	US-09-822-827-848	Sequence 819, App
45	670	39.5	132	9	US-09-895-793-819	Sequence 819, App

ALIGNMENTS

RESULT 1

US-09-886-349A-4
; Sequence 4, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB32A (Ra35 mature)
US-09-886-349A-4

Query Match 100.0%; Score 1698; DB 12; Length 330;

Best Local Similarity 100.0%; Pred. No. 1.1e-121;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MHHHHHAPPALSQDRFADF	PALP	DP	SA	MA	VQ	PV	QV	NT	KL	GY	NA	VG	AG	TG	IV	60
DB	1	MHHHHHAPPALSQDRFADF	PALP	DP	SA	MA	VQ	PV	QV	NT	KL	GY	NA	VG	AG	TG	IV	60
QY	61	PNGVLTNNHVIAGATD	INAF	SV	GS	QY	GV	ND	VT	QV	DR	TD	QV	AV	VL	QL	RAG	120
DB	61	PNGVLTNNHVIAGATD	INAF	SV	GS	QY	GV	ND	VT	QV	DR	TD	QV	AV	VL	QL	RAG	120

QY 121 GVAVGEPVAMNGSGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
 Db 121 GVAVGEPVAMNGSGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
 QY 181 GDSGPPVNVGLGVVGMNTAASDNFQLSQGGQGFPAIPGQAMAIAGQIRSGGSPVTHIG 240
 Db 181 GDSGPPVNVGLGVVGMNTAASDNFQLSQGGQGFPAIPGQAMAIAGQIRSGGSPVTHIG 240
 QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLIGISTGVDVITAVDGPINSATAMADALNGH 300
 Db 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLIGISTGVDVITAVDGPINSATAMADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVT:LAEGPPA 330
 Db 301 HPGDVISVTWQTKSGGTRTGNVT:LAEGPPA 330

RESULT 2

US-10-098-732A-4
 ; Sequence 4, Application US/10098732A
 ; Publication No. US20030175294A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Brannon, Mark
 ; APPLICANT: Guderian, Jeffrey
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
 ; TITLE OF INVENTION: Leishmania Antigen
 ; FILE REFERENCE: 014059-012010US
 ; CURRENT APPLICATION NUMBER: US/10/098,732A
 ; CURRENT FILING DATE: 2003-04-29
 ; PRIOR APPLICATION NUMBER: US 60/275,837
 ; PRIOR FILING DATE: 2001-03-13
 ; NUMBER OF SEQ ID NOS: 80
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 4
 ; LENGTH: 330
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:Ra35 mature
 US-10-098-732A-4

Query Match 100.0%; Score 1698; DB 14; Length 330;
 Best Local Similarity 100.0%; Pred. No. 1.le-121;
 Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVAGTGVIVD 60
 Db 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVAGTGVIVD 60
 QY 61 PNGVLTNNHVIAGATDINAFSVSGGTYGVVDVYDRTQDVAVLQIRGAGGLPSAAIGG 120
 Db 61 PNGVLTNNHVIAGATDINAFSVSGGTYGVVDVYDRTQDVAVLQIRGAGGLPSAAIGG 120
 QY 121 GVAVGEPVAMNGSGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
 Db 121 GVAVGEPVAMNGSGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
 QY 181 GDSGPPVNVGLGVVGMNTAASDNFQLSQGGQGFPAIPGQAMAIAGQIRSGGSPVTHIG 240
 Db 181 GDSGPPVNVGLGVVGMNTAASDNFQLSQGGQGFPAIPGQAMAIAGQIRSGGSPVTHIG 240
 QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLIGISTGVDVITAVDGPINSATAMADALNGH 300
 Db 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLIGISTGVDVITAVDGPINSATAMADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVT:LAEGPPA 330
 Db 301 HPGDVISVTWQTKSGGTRTGNVT:LAEGPPA 330

RESULT 3

US-10-369-983-19
 ; Sequence 19, Application US/10369983
 ; Publication No. US2003023593A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Guderian, Jeff
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009081US
 ; CURRENT APPLICATION NUMBER: US/10/369,983
 ; CURRENT FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: US 60/357,351
 ; PRIOR FILING DATE: 2002-02-15
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 19
 ; LENGTH: 330
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:wild-type
 ; OTHER INFORMATION: mature MTB32A (Ra35)
 US-10-369-983-19

Query Match 100.0%; Score 1698; DB 15; Length 330;
 Best Local Similarity 100.0%; Pred. No. 1.le-121;
 Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVAGTGVIVD 60
 Db 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVAGTGVIVD 60
 QY 61 PNGVLTNNHVIAGATDINAFSVSGGTYGVVDVYDRTQDVAVLQIRGAGGLPSAAIGG 120
 Db 61 PNGVLTNNHVIAGATDINAFSVSGGTYGVVDVYDRTQDVAVLQIRGAGGLPSAAIGG 120
 QY 121 GVAVGEPVAMNGSGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
 Db 121 GVAVGEPVAMNGSGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
 QY 181 GDSGPPVNVGLGVVGMNTAASDNFQLSQGGQGFPAIPGQAMAIAGQIRSGGSPVTHIG 240
 Db 181 GDSGPPVNVGLGVVGMNTAASDNFQLSQGGQGFPAIPGQAMAIAGQIRSGGSPVTHIG 240
 QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLIGISTGVDVITAVDGPINSATAMADALNGH 300
 Db 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLIGISTGVDVITAVDGPINSATAMADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVT:LAEGPPA 330
 Db 301 HPGDVISVTWQTKSGGTRTGNVT:LAEGPPA 330

RESULT 4

US-09-886-349A-6
 ; Sequence 6, Application US/09886349A
 ; Publication No. US20040086523A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009070US
 ; CURRENT APPLICATION NUMBER: US/09/886,349A
 ; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: US 09/597,796
 ; PRIOR FILING DATE: 2000-06-20
 ; PRIOR APPLICATION NUMBER: US 60/265,737
 ; PRIOR FILING DATE: 2001-02-01
 ; NUMBER OF SEQ ID NOS: 50

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 330
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:Ra35FLMutSA
US-09-886-349A-6

Query Match 99.8%; Score 1695; DB 12; Length 330;
Best Local Similarity 99.7%; Pred. No. 1.8e-121;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNVAVGAGTGIVID 60
DB 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNVAVGAGTGIVID 60

QY 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIIG 120
DB 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIIG 120

QY 121 GVAAGPEPVVAMGNSGGQGTTPRAVPGRVVVALGQTVQASDSLTAABETLNLGIQFDAAIOP 180
DB 121 GVAAGPEPVVAMGNSGGQGTTPRAVPGRVVVALGQTVQASDSLTAABETLNLGIQFDAAIOP 180

QY 181 GDSGGPVVNLGQVVGWNTAASDNFQLSQGGQGFALPIQAMAIAGQIRSGGSPVTHIG 240
DB 181 GDSGGPVVNLGQVVGWNTAASDNFQLSQGGQGFALPIQAMAIAGQIRSGGSPVTHIG 240

QY 241 PTAFLGLGVVDNNGARVQVVVGSAPAASLGISTGDTVITAVDGPINSATAMADALNGH 300
DB 241 PTAFLGLGVVDNNGARVQVVVGSAPAASLGISTGDTVITAVDGPINSATAMADALNGH 300

QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 6
US-10-369-983-20
Sequence 20, Application US/10369983
Publication No. US20030235593A1
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Guderian, Jeff
APPLICANT: Reed, Steven
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
FILE REFERENCE: 014058-009081US
CURRENT APPLICATION NUMBER: US/10/369,983
CURRENT FILING DATE: 2003-02-18
PRIOR APPLICATION NUMBER: US 60/357,351
PRIOR FILING DATE: 2002-02-15
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 20
LENGTH: 330
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:mutated
OTHER INFORMATION: MTB32AMutSA (Ra35 mutSA)
US-10-369-983-20

Query Match 99.8%; Score 1695; DB 15; Length 330;
Best Local Similarity 99.7%; Pred. No. 1.8e-121;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNVAVGAGTGIVID 60
DB 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNVAVGAGTGIVID 60

QY 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIIG 120
DB 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIIG 120

QY 121 GVAAGPEPVVAMGNSGGQGTTPRAVPGRVVVALGQTVQASDSLTAABETLNLGIQFDAAIOP 180
DB 121 GVAAGPEPVVAMGNSGGQGTTPRAVPGRVVVALGQTVQASDSLTAABETLNLGIQFDAAIOP 180

QY 181 GDSGGPVVNLGQVVGWNTAASDNFQLSQGGQGFALPIQAMAIAGQIRSGGSPVTHIG 240
DB 181 GDSGGPVVNLGQVVGWNTAASDNFQLSQGGQGFALPIQAMAIAGQIRSGGSPVTHIG 240

QY 241 PTAFLGLGVVDNNGARVQVVVGSAPAASLGISTGDTVITAVDGPINSATAMADALNGH 300
DB 241 PTAFLGLGVVDNNGARVQVVVGSAPAASLGISTGDTVITAVDGPINSATAMADALNGH 300

QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 330
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:Ra35FLMutSA
US-09-886-349A-6

Query Match 99.8%; Score 1695; DB 12; Length 330;
Best Local Similarity 99.7%; Pred. No. 1.8e-121;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNVAVGAGTGIVID 60
DB 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNVAVGAGTGIVID 60

QY 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIIG 120
DB 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIIG 120

QY 121 GVAAGPEPVVAMGNSGGQGTTPRAVPGRVVVALGQTVQASDSLTAABETLNLGIQFDAAIOP 180
DB 121 GVAAGPEPVVAMGNSGGQGTTPRAVPGRVVVALGQTVQASDSLTAABETLNLGIQFDAAIOP 180

QY 181 GDSGGPVVNLGQVVGWNTAASDNFQLSQGGQGFALPIQAMAIAGQIRSGGSPVTHIG 240
DB 181 GDSGGPVVNLGQVVGWNTAASDNFQLSQGGQGFALPIQAMAIAGQIRSGGSPVTHIG 240

QY 241 PTAFLGLGVVDNNGARVQVVVGSAPAASLGISTGDTVITAVDGPINSATAMADALNGH 300
DB 241 PTAFLGLGVVDNNGARVQVVVGSAPAASLGISTGDTVITAVDGPINSATAMADALNGH 300

QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 5
US-10-098-732A-6
Sequence 6, Application US/10098732A
Publication No. US20030175294A1
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Brannon, Mark
APPLICANT: Guderian, Jeffrey
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
FILE REFERENCE: 014058-012010US
CURRENT APPLICATION NUMBER: US/10/098,732A
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: US 60/275,837
PRIOR FILING DATE: 2001-03-13
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 330
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:Ra35FLMutSA
US-10-098-732A-6

Query Match 99.8%; Score 1695; DB 14; Length 330;
Best Local Similarity 99.7%; Pred. No. 1.8e-121;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNVAVGAGTGIVID 60
DB 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNVAVGAGTGIVID 60

QY 61 PNGVLTNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIIG 120

```

RESULT 7
US-10-369-983-2
; Sequence 2, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 723
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: mutated
; OTHER INFORMATION: MTB32-MTB39F fusion protein (MTB32MutSA)
US-10-369-983-2

```

```

Query Match      99.8%; Score 1695; DB 15; Length 723;
Best Local Similarity 99.7%; Pred. No. 4.6e-121;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNAVAGTGIIVD 60
DB 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNAVAGTGIIVD 60
QY 61 PNGVVLNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQRLGAGLPSAAIGG 120
DB 61 PNGVVLNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQRLGAGLPSAAIGG 120
QY 121 GVAVGEPPVAMGNSGGGCTPRVGRVVALGQTVQASDSLTCABETLNGLIQFDDAATOP 180
DB 121 GVAVGEPPVAMGNSGGGCTPRVGRVVALGQTVQASDSLTCABETLNGLIQFDDAATOP 180
QY 181 GDSGPPVNVGLGVVGMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
DB 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
DB 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

```

```

RESULT 8
US-10-369-983-4
; Sequence 4, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22

```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB-102F fusion
US-10-369-983-4

```

```

Query Match      99.8%; Score 1695; DB 15; Length 1010;
Best Local Similarity 99.7%; Pred. No. 7e-121;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNAVAGTGIIVD 60
DB 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNAVAGTGIIVD 60
QY 61 PNGVVLNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQRLGAGLPSAAIGG 120
DB 61 PNGVVLNNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQRLGAGLPSAAIGG 120
QY 121 GVAVGEPPVAMGNSGGGCTPRVGRVVALGQTVQASDSLTCABETLNGLIQFDDAATOP 180
DB 121 GVAVGEPPVAMGNSGGGCTPRVGRVVALGQTVQASDSLTCABETLNGLIQFDDAATOP 180
QY 181 GDSGPPVNVGLGVVGMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
DB 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
DB 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

```

```

RESULT 9
US-09-712-363-161
; Sequence 161, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 161

```

```

; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-161

Query Match
Best Local Similarity 96.4%; Score 1645; DB 9; Length 355;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVNNITKLGYNNAVAGAGTGIVIDPNGVLT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVNNITKLGYNNAVAGAGTGIVIDPNGVLT 92
QY 68 NNHVIAGATDINAFSVGSGQTVGVYVYDRTQDVAVLQRLGAGGLPSAAIGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVGSGQTVGVYVYDRTQDVAVLQRLGAGGLPSAAIGGVAVGEP 152
QY 128 VVAMGNSGGGTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
DB 153 VVAMGNSGGGTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
QY 188 VNGLGQVVGWNTAASDNFOLSGQGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFLGL 247
DB 213 VNGLGQVVGWNTAASDNFOLSGQGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFLGL 272
QY 248 GVVDNNGNGARVQVRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHHPGDVIS 307
DB 273 GVVDNNGNGARVQVRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VTWQTKSGGTRTGNVTLAEGPPA 355

RESULT 10
US-09-886-349a-2
; Sequence 2, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-0090700S
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB32A (Ra35FL)
US-09-886-349A-2

Query Match
Best Local Similarity 99.4%; Score 1634; DB 12; Length 355;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVNNITKLGYNNAVAGAGTGIVIDPNGVLT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVNNITKLGYNNAVAGAGTGIVIDPNGVLT 92
QY 68 NNHVIAGATDINAFSVGSGQTVGVYVYDRTQDVAVLQRLGAGGLPSAAIGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVGSGQTVGVYVYDRTQDVAVLQRLGAGGLPSAAIGGVAVGEP 152
```

```

QY 128 VVAMGNSGGGTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
DB 153 VVAMGNSGGGTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
QY 188 VNGLGQVVGWNTAASDNFOLSGQGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFLGL 247
DB 213 VNGLGQVVGWNTAASDNFOLSGQGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFLGL 272
QY 248 GVVDNNGNGARVQVRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHHPGDVIS 307
DB 273 GVVDNNGNGARVQVRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VTWQTKSGGTRTGNVTLAEGPPA 355

RESULT 11
US-10-193-002-80
; Sequence 80, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campese-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/10/193,002
; APPLICATION NUMBER: US/10/193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 80:
US-10-193-002-80

Query Match
Best Local Similarity 96.2%; Score 1634; DB 14; Length 355;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVVNINTKLGYNNAVAGTGVIDPENGVLVT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVVNINTKLGYNNAVAGTGVIDPENGVLVT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
QY 128 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPGDSGGPV 187
DB 153 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPGDSGGPV 212
QY 188 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPGDSGGPV 247
DB 213 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPGDSGGPV 272
QY 248 GVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 307
DB 273 GVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 12

US-10-084-843-79
; Sequence 79, Application US/10084843
; Publication No. US20030143243A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; Skeiky, Yasir A.W.
; Dillon, Davin C.
; Campes-Neto, Antonio
; Houghton, Raymond
; Wedwick, Thomas S.
; Twardzik, Daniel R.
; Lodes, Michael J.
; Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/084,843
; FILING DATE: 25-Feb-2002
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.

REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

;/ TOPOLOGY: linear
;/ SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-084-843-79

Query Match 96.2%; Score 1634; DB 14; Length 355;
Best Local Similarity 99.4%; Pred. No. 8.9e-117;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVVNINTKLGYNNAVAGTGVIDPENGVLVT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVVNINTKLGYNNAVAGTGVIDPENGVLVT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
QY 128 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPGDSGGPV 187
DB 153 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPGDSGGPV 212
QY 188 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPGDSGGPV 247
DB 213 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPGDSGGPV 272
QY 248 GVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 307
DB 273 GVVDNNGNGARVQVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 13

US-10-098-732A-2
; Sequence 2, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Gudarian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 355
; TYPE: PPT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB32A (Ra35FL)
US-10-098-732A-2

Query Match 96.2%; Score 1634; DB 14; Length 355;
Best Local Similarity 99.4%; Pred. No. 8.9e-117;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVVNINTKLGYNNAVAGTGVIDPENGVLVT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVVNINTKLGYNNAVAGTGVIDPENGVLVT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
QY 128 VVAMGNSGGGQGTTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPGDSGGPV 187

Db 153 VVAMGNSGGGTPRAVPGRVVALGQTQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 212
QY 188 VNLGQVVGWNTAASDNFQLSQGGQGFALPIQCAAMAIAGQIRSGGSPVTHIGPTAFGL 247
Db 213 VNLGQVVGWNTAASDNFQLSQGGQGFALPIQCAAMAIAGQIRSGGSPVTHIGPTAFGL 272
QY 248 GVVDNNGNGARVQVVGSAASLIGTGDVITAVDVGAPINSATAMADALNGHHFGDVIS 307
Db 273 GVVDNNGNGARVQVVGSAASLIGTGDVITAVDVGAPINSATAMADALNGHHFGDVIS 332
QY 308 VVWTKSGGTRGTCNVTLAEGPPA 330
Db 333 VVWTKSGGTRGTCNVTLAEGPPA 355

RESULT 14
US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85b)
US-10-369-983-18

Query Match 59.1%; Score 1003.5; DB 15; Length 1016;
Best Local Similarity 85.9%; Pred. No. 4.6e-68;
Matches 207; Conservative 6; Mismatches 13; Indels 15; Gaps 4;
QY 8 APPALSQDRPADFPALPLDPSAMVAQVGPQVNVNITKLGYNNAVAGTGVIDPNGVILT 67
Db 535 APPALSQDRPADFPALPLDPSAMVAQVGPQVNVNITKLGYNNAVAGTGVIDPNGVILT 594
QY 68 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVILQIRGAGGLPSAAIGGGVAVGEP 127
Db 595 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVILQIRGAGGLPSAAIGGGVAVGEP 654
QY 128 VVAMGNSGGGTPRAVPGRVVALGQTQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 187
Db 655 VVAMGNSGGGTPRAVPGRVVALGQTQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 714
QY 188 VNLGQVVGWNTAASDNFQLSQGGQGFALPIQCAAMAIAGQIRSGGSPVTHIGPTAFGL 238
Db 715 VNLGQVVGWNTAASGTF--SRPG-----LPEVYLQVPSMGRDVKVQFGSGGNNSPAVY 768
QY 239 I 239
Db 769 L 769

RESULT 15
US-10-369-983-16
; Sequence 16, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff

; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein
; OTHER INFORMATION: MTB114F (MTB72F-mTCC#2)
US-10-369-983-16

Query Match 58.6%; Score 995; DB 15; Length 1154;
Best Local Similarity 73.2%; Pred. No. 2.4e-67;
Matches 213; Conservative 13; Mismatches 37; Indels 28; Gaps 4;
QY 8 APPALSQDRPADFPALPLDPSAMVAQVGPQVNVNITKLGYNNAVAGTGVIDPNGVILT 67
Db 535 APPALSQDRPADFPALPLDPSAMVAQVGPQVNVNITKLGYNNAVAGTGVIDPNGVILT 594
QY 68 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVILQIRGAGGLPSAAIGGGVAVGEP 127
Db 595 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVILQIRGAGGLPSAAIGGGVAVGEP 654
QY 128 VVAMGNSGGGTPRAVPGRVVALGQTQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 187
Db 655 VVAMGNSGGGTPRAVPGRVVALGQTQASDSLTGAETLNGLIQFDAAIQPGDSGGPV 714
QY 188 VNLGQVVGWNTAASDNFQLSQGGQGFALPIQCAAMAIAGQIRSGGSPVTHIGPTAFGL 247
Db 715 VNLGQVVGWNTAASGTFMDF-----GLLPP-----EVNSSRMYSGPSPESMLAAAAWDGV 765
QY 248 GVVDNNGNGARVQVVGSAASLIGTGDVITAVDVGAPI--NSATAMADA 296
Db 766 -----AAELTSAAVSYGSVSTLIVEPMWMPAAAAA 799

Search completed: June 22, 2004, 18:07:52
Job time : 31.5644 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:59:04 ; Search time 9.233 Seconds
(without alignments)
3438.018 Million cell updates/sec

Title: US-09-886-349A-4
Perfect score: 1698
Sequence: 1 MHHHHHAPPALSDQRFADP.....QTKSGGTRGNVTLAEGPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78: **
1: pir1: **
2: pir2: **
3: pir3: **
4: pir4: **

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1645	96.9	355	2 F70983	probable serine pr
2	1231.5	72.5	361	2 S47170	hypothetical prote
3	1198	70.6	354	2 A87242	probable secreted
4	459.5	27.1	464	2 C70821	probable serine pr
5	450	26.5	382	2 H86930	probable secreted
6	450	26.5	452	2 T45448	probable serine pr
7	374	22.0	394	2 S74643	proteinase hhoA (E
8	368	21.7	407	2 AG2150	serine proteinase
9	333	20.8	433	2 H97199	htrA-like serine p
10	332	20.7	452	2 S77538	serine proteinase
11	338	19.9	514	2 A82581	periplasmic protei
12	335.5	19.8	362	2 T35287	probable secreted
13	333	19.6	416	2 AB2057	serine proteinase
14	332	19.6	398	2 B71284	probable periplasm
15	332	19.6	441	2 E75357	probable periplasm
16	329.5	19.4	429	2 AD1894	serine proteinase
17	319.5	18.8	474	2 F83550	serine proteinase
18	319	18.8	355	1 JC6052	trypsin-like prote
19	319	18.8	355	2 D91142	trypsin-like prote
20	318	18.7	355	2 G85987	trypsin-like prote
21	317.5	18.7	401	2 AD2451	serine proteinase
22	315	18.6	408	2 H86891	exported serine pr
23	313.5	18.5	352	2 E82307	trypsin-like prote
24	312.5	18.4	455	2 AB0909	serine protease (E
25	312	18.4	416	2 S75445	proteinase hhoB (E
26	310.5	18.3	348	2 H96956	serine protease Do
27	310.5	18.3	455	2 C91142	serine endoprotein
28	309	18.2	455	2 F85987	serine endoprotein
29	308.5	18.2	513	2 AD3418	proteinase DO (EC

30	307.5	18.1	513	2	I40060	serine proteinase
31	306.5	18.1	474	1	I40059	htrA-like protein
32	306.5	18.1	474	2	A13349	proteinase DO (EC
33	305.5	18.0	455	2	JC6051	trypsin-like prote
34	304	17.9	328	1	D69109	serine proteinase
35	302.5	17.8	356	2	AC0909	trypsin-like prote
36	301	17.7	459	2	F72359	periplasmic serine
37	299	17.6	457	2	AG0433	proteinase (EC 3.4
38	298	17.6	499	2	H81914	probable periplasm
39	296	17.4	475	1	S15337	heat shock protein
40	294.5	17.3	363	2	C87336	serine proteinase
41	294.5	17.3	449	2	A69643	serine proteinase
42	294.5	17.3	474	2	S45229	proteinase DO (EC
43	294.5	17.3	474	2	S85500	proteinase DO (EC
44	294.5	17.3	474	2	E90649	proteinase DO (EC
45	294	17.3	476	2	H71936	proteinase DO - He

ALIGNMENTS

RESULT 1

F70983
probable serine proteinase: pepA - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002
C:Accession: F70983
R:Coile, S.I.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: F70983
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-355 <COL>
A:Cross-references: GB:Z96071; GB:AL123456; NID:g3242254; PIDN:CAB09453.1; PID:g2181967
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: pepA
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; trypsin

Query Match	96.9%	Score 1645;	DB 2;	Length 355;
Best Local Similarity	100.0%;	Pred. No. 9.4e-90;		
Matches 323;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	8	APPALSDQRFADFPALPLDPSAMVAQVGPQVYNTKLGYNNAVAGTGVIDPENGVLVT	67	
Db	33	APPALSDQRFADFPALPLDPSAMVAQVGPQVYNTKLGYNNAVAGTGVIDPENGVLVT	92	
QY	68	NNHVIAGATDINAFSVGSGQTVGVYDRTQDVAVIQLRGAGLPSAALGGGVAVGCEP	127	
Db	93	NNHVIAGATDINAFSVGSGQTVGVYDRTQDVAVIQLRGAGLPSAALGGGVAVGCEP	152	
QY	128	VVMGNSGGGSGTTPRAVGRVVALQQTVAQSDSLTGAEETLNGLIQFDAAIQPDGSGGPV	187	
Db	153	VVMGNSGGGSGTTPRAVGRVVALQQTVAQSDSLTGAEETLNGLIQFDAAIQPDGSGGPV	212	
QY	188	VNGLGVQVGMNTAASDNFQLSQGGGGAIPGQAWATAGQIRSGGSGTTHIGTAFGL	247	
Db	213	VNGLGVQVGMNTAASDNFQLSQGGGGAIPGQAWATAGQIRSGGSGTTHIGTAFGL	272	
QY	248	GVVDNNGNGARVQRVVGSAASLGISTGDIVITAVDGAPINSATAMADALNGHPGDVIS	307	
Db	273	GVVDNNGNGARVQRVVGSAASLGISTGDIVITAVDGAPINSATAMADALNGHPGDVIS	332	
QY	308	VTWOTKSGGTRTGNVTLAEGPPA	330	
Db	333	VTWOTKSGGTRTGNVTLAEGPPA	355	

Db 436 APGATVALTPQDPGSGSRTVQVTLGK 461

RESULT 5

H86930

Probable secreted serine proteinase [imported] - Mycobacterium leprae

C;Species: Mycobacterium leprae

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002

C;Accession: H86930

R;Cole, S.F.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho

eam, M.A.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltham, T.; Fraser, A.; Hamlin, N.; Holroyd,

Nature 409, 1007-1011, 2001

A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S

A;Title: Massive gene decay in the leprosy bacillus.

A;Reference number: A86909; MUID:21128732; PMID:11234002

A;Accession: H86930

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-382 <STO>

A;Cross-references: GB:AL450380; NID:gl13092536; PIDN:CAC29684.1; GSPDB:GNC0147

C;Genetics:

A;Gene: M0176

C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Query Match 26.5%; Score 450; DB 2; Length 382;

Best Local Similarity 37.5%; Pred. No. 1.4e-19;

Matches 126; Conservative 52; Mismatches 116; Indels 42; Gaps 12;

QY 18 ADFPALPLDPSAMVAQVG-----PQVNNINTKLGNNAVAGAGTIVDPNGVLTNNHVA 73

Db 59 ASVPAANM-PSGSVEQAVKVPVSVNLETLGRQSE--EGSGVILSADGLILTNHVA 115

QY 74 GA-----TDINAFVSGGQTYGVVDVGRDTPQVAVLQLRGAGGLPSAAIG 119

Db 116 VAAKPGGGGGLSPKTTVTFF--DGRASFTVVGADPTSDIAVVRVQSISGLTPTMG 172

QY 120 GG--VAVGEPPVAMGNSGGGGTTPRAVGRVVALGTQVQASDLSLTGAETLNGLIQFDAA 177

Db 173 SSADLRVQGVVAVGSPGLAGT--VTSGIVSALNRPVSTTGE--SGNQTIVLDAIQTDAA 229

QY 178 IQPDSGGPPVNGLGQVGVGNMTAA-----SDNFQLSQGGQFAIPGQAVIAGQIRSG 231

Db 230 INFGSGGALVNMGGQLVGVNSAIATLGADSGAQSGSIGLGFPAIPVDQAKRIADELIST 289

QY 232 GGSPTVHIGPTAFGLGVVDNNGN--GARVQVVGSAASLSLTGDTAVDVGAPINSA 290

Db 290 G--KATH---ASLGVQVATDKGTGPAKVMVDVAGGAANAAPKGVVLTQVDDRLISSA 343

QY 291 TAMADALNGHHPGDVSVTWQTSKGGTRTGNVTLAE 326

Db 344 DALVAARSKAPGDKVSLTYQDSGSSRTVQVTLGK 379

RESULT 6

T45448

Probable serine proteinase (EC 3.4.21.-) MLCB373.28 [similarity] - Mycobacterium leprae

C;Species: Mycobacterium leprae

C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Dec-2002

C;Accession: T45448

R;James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, February 1998

A;Reference number: Z22967

A;Accession: T45448

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-452 <JAM>

A;Cross-references: EMBL:AL035500; PIDN:CAB36690.1

A;Experimental source: cosmid L373

C;Genetics:

A;Note: MLCB373.28

C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

C;Keywords: hydrolase; serine proteinase

Query Match 26.5%; Score 450; DB 2; Length 382;

Best Local Similarity 37.5%; Pred. No. 1.4e-19;

Matches 126; Conservative 52; Mismatches 116; Indels 42; Gaps 12;

QY 18 ADFPALPLDPSAMVAQVG-----PQVNNINTKLGNNAVAGAGTIVDPNGVLTNNHVA 73

Db 59 ASVPAANM-PSGSVEQAVKVPVSVNLETLGRQSE--EGSGVILSADGLILTNHVA 115

QY 74 GA-----TDINAFVSGGQTYGVVDVGRDTPQVAVLQLRGAGGLPSAAIG 119

Db 116 VAAKPGGGGGLSPKTTVTFF--DGRASFTVVGADPTSDIAVVRVQSISGLTPTMG 172

QY 120 GG--VAVGEPPVAMGNSGGGGTTPRAVGRVVALGTQVQASDLSLTGAETLNGLIQFDAA 177

Db 173 SSADLRVQGVVAVGSPGLAGT--VTSGIVSALNRPVSTTGE--SGNQTIVLDAIQTDAA 229

QY 178 IQPDSGGPPVNGLGQVGVGNMTAA-----SDNFQLSQGGQFAIPGQAVIAGQIRSG 231

Db 230 INFGSGGALVNMGGQLVGVNSAIATLGADSGAQSGSIGLGFPAIPVDQAKRIADELIST 289

QY 232 GGSPTVHIGPTAFGLGVVDNNGN--GARVQVVGSAASLSLTGDTAVDVGAPINSA 290

Db 290 G--KATH---ASLGVQVATDKGTGPAKVMVDVAGGAANAAPKGVVLTQVDDRLISSA 343

QY 291 TAMADALNGHHPGDVSVTWQTSKGGTRTGNVTLAE 326

Db 344 DALVAARSKAPGDKVSLTYQDSGSSRTVQVTLGK 379

F.182,224,305/Active site: His, Asp, Ser #status predicted

Query Match 26.5%; Score 450; DB 2; Length 452;

Best Local Similarity 37.5%; Pred. No. 1.7e-19;

Matches 126; Conservative 52; Mismatches 116; Indels 42; Gaps 12;

QY 18 ADFPALPLDPSAMVAQVG-----PQVNNINTKLGNNAVAGAGTIVDPNGVLTNNHVA 73

Db 129 ASVPAANM-PSGSVEQAVKVPVSVNLETLGRQSE--EGSGVILSADGLILTNHVA 185

QY 74 GA-----TDINAFVSGGQTYGVVDVGRDTPQVAVLQLRGAGGLPSAAIG 119

Db 186 VAAKPGGGGGLSPKTTVTFF--DGRASFTVVGADPTSDIAVVRVQSISGLTPTMG 242

QY 120 GG--VAVGEPPVAMGNSGGGGTTPRAVGRVVALGTQVQASDLSLTGAETLNGLIQFDAA 177

Db 243 SSADLRVQGVVAVGSPGLAGT--VTSGIVSALNRPVSTTGE--SGNQTIVLDAIQTDAA 299

QY 178 IQPDSGGPPVNGLGQVGVGNMTAA-----SDNFQLSQGGQFAIPGQAVIAGQIRSG 231

Db 300 INFGSGGALVNMGGQLVGVNSAIATLGADSGAQSGSIGLGFPAIPVDQAKRIADELIST 359

QY 232 GGSPTVHIGPTAFGLGVVDNNGN--GARVQVVGSAASLSLTGDTAVDVGAPINSA 290

Db 360 G--KATH---ASLGVQVATDKGTGPAKVMVDVAGGAANAAPKGVVLTQVDDRLISSA 413

QY 291 TAMADALNGHHPGDVSVTWQTSKGGTRTGNVTLAE 326

Db 414 DALVAARSKAPGDKVSLTYQDSGSSRTVQVTLGK 449

RESULT 7

S74643

Proteinase rhoA (EC 3.4.-.-) - Synecocystis sp. (strain PCC 6803)

N;Alternate names: protein sll1679

C;Species: Synecocystis sp.

A;Variety: PCC 6803

C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Dec-2002

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;

O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,

DNA Res. 3, 109-136, 1996

A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis

S.

A;Reference number: S74322; MUID:97061201; PMID:8905231

A;Accession: S74643

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-394 <KAN>

A;Cross-references: EMBL:D90900; GB:AB001339; NID:gl651768; PIDN:BAAL6795.1; PID:gl651861

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C;Genetics:

A;Gene: rhoA

C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

C;Keywords: hydrolase; proteinase

Query Match 22.0%; Score 374; DB 2; Length 394;

Best Local Similarity 31.6%; Pred. No. 4.3e-15;

Matches 118; Conservative 49; Mismatches 111; Indels 96; Gaps 15;

QY 9 PPA-----LSQDRFADFAPALPLDPSAMVAQVQGVVNTK-----44

Db 39 PPAPVITAQASVPLTSESFV-----AAVSRSGEAVVDTETVTRTRDPIIDDP 89

QY 45 -----LGNNVAV-----GAGTGIVDPNGVLTNNHVIAGATDINAFVSGSQTYGV 91

Db 90 FQEFGRSFPVPRERRIAGGSGFIIDNSGIILTAHVVDGASKV-VVTLRDGRITDG 148

QY 92 DVVGYDRTQDVAVLQLRGAG-GLPSAAIG--GGVAVGEPPVAMGNSGGGGTTPRAVGRV 148

Db 149 QVRGDEVTDLAVAVIEPQGGALPVAPLGTSSNQLQVGDWAIAGNPFVGLDNT-----200

QY 149 VALG--QTQVQASDLSLTGAETLNGLIQFDAAIQPDSGGPVVNGLGQVGMNTAASDNFQ 206

Db 201 VTLLGIITSLGSAQAQIPKRVFEIQTDAINPNSGGPPLLARGVEIGINTA---IR 256
 Qy 207 LSQGGQGFAPIGQMAIAGQIRSGGSPVTHIGPTAFGLGLGV---VDNNGN----- 255
 Db 257 ADATGIGFAIPIDQAKAIONTLAAGTVPHPIYG-VQMMNITVDQAQNNRNPNFPPIIP 315
 Qy 256 ---GARVQVVVGSAPAASLIGSTGSDVITAVDGAPINSATAM-----ADALN 298
 Db 316 EVDGILVMRLPGTPAERAGIRRGDVIIVAVDGTPIISDGAQLQRIVEQAGLKLKLLR 375
 Qy 299 GHFPGDVISVTWQT 312
 Db 376 GDRR---LSLTVQT 386

RESULT 8
 AG2150
 serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AG2150
 R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
 DNA Res. 6, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AG2150
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-407 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA574457.1; PID:g17131851; GSPDB:GNC00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr2758
 C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp

Query Match 21.7%; Score 368; DB 2; Length 407;
 Best Local Similarity 31.5%; Pred. No. 1e-14;
 Matches 110; Conservative 51; Mismatches 118; Indels 70; Gaps 11;
 Qy 28 SAMVAQVGPQVNNITKLGNNAV-----GAGTGIVID 60
 Db 74 TAAVNRVGSAAVRIDTERTITRRVDPELPPFRFRFEGFGQQLPPEQMRGLSGFIID 133
 Qy 61 PNGVLTNNHVIAGATDINAFSVSGQTYGVVDVGYDRTODVAVLQLRGAGGLPSAAIG- 119
 Db 134 KSGILTNHVVDAKDRTV-RLKXGSRFGKQGVDEVTDLAVVKLNAGNSLPVAPLGS 192
 Qy 120 -GGVAGPEPVVAMNSGGQGGTTPRAVGRVVALG--QTVAQSDSLTGAETLNLQFDA 176
 Db 193 SNNVQVGDWALAVGNPLGFONT-----VTLGIVSTLKRSSAQVGIITDKRLDFIQTDA 244
 Qy 177 AIQPDGSGPVVNLGLGVQVGMNTAASDNFQLSQGGQGFAPIGQMAIAGQIRSGGSP 236
 Db 245 AINPNSGGPPLNDKGEVIGINTA-----IRADAMGIGFAIPIDKAKAIAITQLERDG--KV 298
 Qy 237 VHIGPTAFGLGV-----VDNNG-----NGARVQVVGSAPAASLIGSTGSD 277
 Db 299 AH-----PYLGVQMATLPFLAQNNIDPNSAFAPPEVGVVLRVSVNSPAAANAGIRRGD 354
 Qy 278 VITAVDGAPINSATAMADALNHHPGDVIVSWTQKSGGTRTGNVILAE 326
 Db 355 VILQVDGQAITTAQLQNVVNSRLGQALQVRLQ-RGNQTQLSLVRTAE 402

RESULT 9
 H97199
 htra-like serine protease (with PDZ domain) [imported] - Clostridium acetobutylicum
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Dec-2002
 C:Accession: H97199

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: H97199
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-433 <KUR>
 A:Cross-references: GB:AB001437; PIDN:AAK80387.1; PID:g15025449; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC2433
 C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp

Query Match 20.8%; Score 353; DB 2; Length 433;
 Best Local Similarity 32.0%; Pred. No. 8.2e-14;
 Matches 101; Conservative 50; Mismatches 111; Indels 54; Gaps 10;
 Qy 25 LDPSAMVACVGPQVNNITKL-----GYNNVAGAGTGVIDPENGVLVT 67
 Db 122 LTVSIVVKKVSPAVVGVSTKTTVTQNDPDSFPGSSNGSGSTQEGMGGIIFNNDGYLT 181
 Qy 68 NNHVIAGATDINAFSVSGQTYGVVDVGYDRTQDVAVLQLRGAGGLPSAAIGGVA---V 124
 Db 182 NYHVIKGAADKI-AVILNNKKEVSARKVNVYDEANDIAVTKMTGSFTVPGVAELGSSASLNV 240
 Qy 125 GPVVMNSGGQGGTTPRAVGRVVALGQTYQASDSLTAETLNLQFDAAIQPDGSG 184
 Db 241 GDSVVAIGNPLGKEPLGLVTTGVSVNRVAVSE---GQKQT---YIQTDAINPNSG 294
 Qy 185 GPVNVNLGQVGMNTAASDNFQLSQGGQ---GPAIPIGQMAIAGQIRSGGSPVTHIGP 241
 Db 295 GPLVNSFGQVVGINS-----KISENGVEGIGFIP-DTVKSKIQNLK-----P 339
 Qy 242 TAFGLG-VVD-----NNGNGARVQVVGSAPAASLIGSTGSDVITAVDGAPINSATA 292
 Db 340 IIMLGIGSEAVDKSPAEGHNPQGVYIEIQDFSSAQKAGMQVGDVITKFDGKKVTSTSD 399
 Qy 293 MADALNGHHPGDVIVS 308
 Db 400 IDSIXSKNSGDTVOV 415

RESULT 10
 S77538
 serine proteinase (EC 3.4.21.-) htra - Synechocystis sp. (strain PCC 6803)
 N:Alternate names: protein slr1204
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Dec-2002
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S77538
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-452 <KAN>
 A:Cross-references: EMBL:D90905; GB:AB001339; NID:g1652360; PIDN:BAAL17385.1; PID:g165246;
 A:Note: The nucleotide sequence was submitted to the ENEL Data Library, June 1996
 C:Genetics:
 A:Gene: htra
 C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp
 C:Keywords: hydrolase; serine proteinase

Query Match 20.7%; Score 352; DB 2; Length 452;
 Best Local Similarity 30.2%; Pred. No. 9.8e-14;
 Matches 110; Conservative 59; Mismatches 113; Indels 82; Gaps 15;
 Qy 28 SAMVAQVGPQVNNITKLGNNAV-----GAGTGIVID 60
 Db 74 TAAVNRVGSAAVRIDTERTITRRVDPELPPFRFRFEGFGQQLPPEQMRGLSGFIID 133
 Qy 61 PNGVLTNNHVIAGATDINAFSVSGQTYGVVDVGYDRTODVAVLQLRGAGGLPSAAIG- 119
 Db 134 KSGILTNHVVDAKDRTV-RLKXGSRFGKQGVDEVTDLAVVKLNAGNSLPVAPLGS 192
 Qy 120 -GGVAGPEPVVAMNSGGQGGTTPRAVGRVVALG--QTVAQSDSLTGAETLNLQFDA 176
 Db 193 SNNVQVGDWALAVGNPLGFONT-----VTLGIVSTLKRSSAQVGIITDKRLDFIQTDA 244
 Qy 177 AIQPDGSGPVVNLGLGVQVGMNTAASDNFQLSQGGQGFAPIGQMAIAGQIRSGGSP 236
 Db 245 AINPNSGGPPLNDKGEVIGINTA-----IRADAMGIGFAIPIDKAKAIAITQLERDG--KV 298
 Qy 237 VHIGPTAFGLGV-----VDNNG-----NGARVQVVGSAPAASLIGSTGSD 277
 Db 299 AH-----PYLGVQMATLPFLAQNNIDPNSAFAPPEVGVVLRVSVNSPAAANAGIRRGD 354
 Qy 278 VITAVDGAPINSATAMADALNHHPGDVIVSWTQKSGGTRTGNVILAE 326
 Db 355 VILQVDGQAITTAQLQNVVNSRLGQALQVRLQ-RGNQTQLSLVRTAE 402

QY 24 PLDPAMVAQV---GPOVNTNKLGVNNAV-----G 52
 DB 112 PREPSNFVVDVSTGPAVVRINAQTKVSKQVPOAFNDPFLQRFSGQMPMPNERNVORG 171
 QY 53 ACTGVVDGNGVVLNNHVIAGATDINAFSGSGGTGYVDVGYDRTODVAVLQRLGAGG 112
 DB 172 TSGGFLVSDNGKIFNAHVDGADVVV-TLXDRSFRGVRVSGSPSTDAVVKIE-AGD 229
 QY 113 LPSAAIGGG--VAVGEPVVMNGSGGGGTTPRAVGRVVALGQTVQASDLSLTGABEETUNG 170
 DB 230 LPTVALGDSHDLQVGEWALAIQNLGLDNT--VTTGILSATGR--RSAD--IGVDPKREVE 283
 QY 171 LTFDAAIOPGSGGVPVNGLVGVGMNTAASDNFQLSGOGGFAIPIGOAMAIAGQIRS 230
 DB 284 FIQTDAINPGSGGGLLADGGVIMNTAIONAQ-----GIGPALPINKAQEIIAQOLIA 339
 QY 231 GGGSPVTHIGPTAFGLGLGVVDNN-----GNGARVORVVGSAAPASLGI 273
 DB 340 TG--KVEH---AYLGIQVMTVPELQSOIRQETGMNI PVDKGVIMQVMPNSPAAIAKL 393
 QY 274 STGDVITAVDGPINATAMADALNGHHPGDVIVTWQTKSGGTRTG---NVTLAGG--- 327
 DB 394 EQGDVLQSLQGPVNAEQVQSLVGLAVGDEVEL-----GILRNGQOONLTVTIGALP 447
 QY 328 --PP 329
 DB 448 SAPP 451

 RESULT 11
 A:Gene: Xylella fastidiosa periplasmic proteinase XF2241 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: A82581
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
 A:Title: Nature 406, 151-157, 2000
 A:Reference number: A82515; PMID:20355717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: A82581
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-514 <SIM>
 A:Cross-references: GB:AE003849; NID:g9107394; PIDN:AAF85040.1; GSPDB:GN001
 A:Experimental source: Strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H. as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S. submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.B.; Laigh Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF2241
 C:Superfamily: Helicobacter serine proteinase

 Query Match 19.9%; Score 338; DB 2; Length 514;
 Best Local Similarity 32.6%; Pred. No. 7.5e-13;
 Matches 109; Conservative 48; Mismatches 107; Indels 70; Gaps 13;

 QY 23 LPLDPSAMVAQVGPQVNNINT-----KLGVNNAV----- 51
 DB 54 LP-DFTQLVDQVGPVNNIETVTRKKVGRGIPLNDIPEFPRFRFGPDMQNPQRG 112
 QY 52 -----GAGTGIVDNGVVLNNHVIAGATDINAFSGSGGTGYVDVVGDRTO 100

DB 113 QODDEGGIAGGNGSGFLISKDGVILNNHVIAGSEV-TIKLTDREFFRAKIIGSEQY 171
 QY 101 DVAVLQRLGAGGIPSAIAGGVAV--GEPVVMNGSGGGGTTPRAVGRVVALGQTVQAS 158
 DB 172 DVALLKI-DARKNLPTRVIGDSSSLKSGQVVAIGSPGLDHSVTA--GIVSALGRS----- 224
 QY 159 DSLTGAREETLNGLIQFDAAIOPGDSGGVPVNGLVGVGMNTAASDNFQLSQG--GQGFAL 216
 DB 225 ---TSDQRYVPFLOTDPINQNGSGGPLNTRGEVIGIN---SQIFASGGYMGISFAI 278
 QY 217 PIGQAMAIAGQIRSGG---GSPVTHIGPTAFGL---GLGVVDNNGCARVORVVGSAAP 269
 DB 279 PINLAINAAEQIRKTKGVQSRMLGVEIGPIDALKAQGLGLPDS--RGALVNNIIPHPSPA 336
 QY 270 SLGISTGDVITAVDGPINATAMADALNGHHPG 303
 DB 337 KAGIEVGDVIRSVNGKVSISSFDLPPLIGMPPG 370

 RESULT 12
 T35287
 probable secreted proteinase - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Dec-2002
 C:Accession: T35287
 R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, July 1999
 A:Reference number: Z21574
 A:Accession: T35287
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-362 <SES>
 A:Cross-references: EMBL:AL096872; PIDN:CAB51255.1; GSPDB:GN00070; SCODEB:SCSF7.30
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCODEB:SCSF7.30
 C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryps

 Query Match 19.8%; Score 335.5; DB 2; Length 362;
 Best Local Similarity 33.9%; Pred. No. 7.2e-13;
 Matches 113; Conservative 45; Mismatches 138; Indels 37; Gaps 12;

 QY 8 APPALSDQRFADFPALPLDPSAMVAQVGPQVNNIKNLVNNAVCGAGTGIVDPNGVLT 67
 DB 52 APRAASE-----LEADYERVIKDVLPVSVVQIA-----GDSLGSVVVDKGHVVT 97
 QY 68 NNHVIAGATDINAFSGSGGTGYG---VDVWGYDRTQDVAVLQL-RGAGGLPSA--AIGGG 121
 DB 98 NARVVG---DAKSFVTTARTEGALTAKLVSSYPQDLAVIKLDKVPGRMARAFADSAK 154
 QY 122 VAVGSPVVMNGSGGGGTTPRAVGRVVALGQTVQASDLSLTGABEETLNGLIQFDAAIOPG 181
 DB 155 VEVGQIVLAMSGPLGLSSS--VTQIVSATGRTVTEGSSGGGTGATIANMVQTSAAINPG 212
 QY 182 DSGGVPVNGLVGVGMNTAASDNFQLSQG---GQGFALPIGQAMAIAGQIRSGG---GSP 235
 DB 213 NSGGALVNDGGQVIGIPILAIATDPGLGSAAPGIGFALPASMTVTVAGQIVRDKVTDG 272
 QY 236 TVHIGPTAFGLGVVDNNGN--GARVORVVGSAAPASLGISTDVITAVDGPINATAM 293
 DB 273 RAALGITA---RTVWDDSVRPAGAAVVEVSDGGAADAGLRPGDVLVGLGIDTITITSL 329
 QY 294 ADALNGHHPGDVIVTWQTKSGGTGTGNVTLAE 326
 DB 330 SEALASMRPGDRKVTY-TRDGKHTAEVTLGE 361

 RESULT 13
 AB2057
 serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C/Accession: AB2057
 R:Kaneke, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
 DNA Res. 8, 205-213, 2001
 A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
 A/Reference number: AB1807; MUID:21595285; PMID:11759840
 C/Accession: AB2057
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-416 <KUR>
 A/Cross-references: GB:BA000019; PIDN:BA073707.1; PID:g17131098; GSPDB:GN00179
 A/Experimental source: strain PCC 7120
 C/Genetics:
 A/Gene: all2008
 C/Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; try

Query Match 19.6%; Score 333; DB 2; Length 416;
 Best Local Similarity 29.0%; Pred. No. 1.2e-12;
 Matches 101; Conservative 63; Mismatches 126; Indels 58; Gaps 12;

QY 26' DPS---AMVAQGPQVNNITKLGYNNAV-----GAGT 55
 DB 77 DPNFVGWVQVGVAVRIDSARTVTSRVPNEFPFRFFGFGVPAQPRQVRERGS 136

QY 56 GIVIDPNCVLTNNHVIAGATDINAFVSGGTQGVVDVVDRTQDVAVQLRAGGLPS 115
 DB 137 GFIISSGQILTNHNVGDADEVTV-TLKDRSGDGVKLGEDPVDVAVIQI-NANNLEP 194

QY 116 AAGGGVAV--GEPVWAMNGSGGGPRAVPRGVVAVGQVQASDLSLTCAEETLGLIQ 173
 DB 195 VAVGNSEVLQGEAVIAIGNPLGNLS--VTSGLIATGR----SGSDIGASRKYVDYIQ 248

QY 174 FDAIQQGSGGPPVNGLGQVGVGNNTAASDNFQSQGGQFAIPIGQAMAIAQIRSGG 233
 DB 249 TDAINFNGSGGGLNARGGVIGMNTAIIQGAQ---GLGFAIPINTVQKVSQELITQK 304

QY 234 SPTVHIG-PTAFGLGVVD--NNGGARVQ-----RVVGSAPASLSTGCVIT 280
 DB 305 VDHPYGVQVATLPQVKERINERFGRINITADRGVLLVIVPGSPAANAGLRPGDIQ 364

QY 281 AVTGAPINSATAMADALNGHHPGDEVISVTWQTSKGGTRTGNVTLAERG 328
 DB 365 SINQSVTTVEQVKIVENSQIGQPLQIE-RNGQTTQVNVSPALP 411

RESULT 14
 B71284
 probable periplasmic serine proteinase DO (htrA-1) - syphilis spirochete
 C/Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
 C/Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 09-Dec-2002
 C/Accession: B71284
 R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
 rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
 they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
 Science 281, 375-388, 1998
 A/Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
 A/Reference number: B71250; MUID:98332770; PMID:9665876
 C/Accession: B71284
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-398 <COL>
 A/Cross-references: GB:AE001248; GB:AE000520; NID:g3323074; PIDN:AAC65740.1; PID:g332308
 A/Experimental source: strain Nichols
 C/Genetics:
 A/Gene: TP0773
 C/Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; try

Query Match 19.6%; Score 332; DB 2; Length 398;
 Best Local Similarity 32.4%; Pred. No. 1.3e-12;
 Matches 104; Conservative 42; Mismatches 121; Indels 54; Gaps 11;

DB 72 AVYRSANEAVNITTEMVGNWFLFPVPLEGGSGGAIIDARGVLTNTHTVEGASKI-Y 130
 QY 81 FSVGSGQTQGVVDVVDRTQDVAVQLRAGAGLPSAAIGGG---VAVGEPVWAMNGSG 136
 DB 131 LSLHDSQYKATVGVVDRENDLAVLKFSVPPGARLTVIRFGSSRLNDVGQKVLAIQNPFG 190
 QY 137 QGGTPRAVPCRVVALGQTVQASDLSLTGAEETLGLIQDAAIQPGDGGSGPVVNGAGVVG 196
 DB 191 LART--LTGVVVSALARPIQNKGS-----IRNMIQTDALNPNNGSGPLDQGRMIG 242
 QY 197 MNTAASDNFQSQGGQFAIPIGQAMAIA-----GQIRSG---CGSPVTHIGPTAFGL 247
 DB 243 INTVIYST-SGSSSGVGFAVPVDTAKRIVSELIRYGRVRRGKIDAEVLQVNVASTAHYAQL 301
 QY 248 GVVDDNNGARVQVYVVGAPASLGLS-----TGCVITAVDVGAPIN 288
 DB 302 TV---GKLLVSVQVRGSPAAQAGLRGTTAVRYGLGRRAAVIYVVGVDVITADNQPVA 357
 QY 289 SATAMADALNGHHPGDEVISVT 309
 DB 358 NLSDYVSVLEDKKPDDEVRT 378

RESULT 15
 E75357
 probable periplasmic serine proteinase Do - Deinococcus radiodurans (strain R1)
 C/Species: Deinococcus radiodurans
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Dec-2002
 C/Accession: E75357
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; I
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A/Reference number: A75250; MUID:20036896; PMID:10567266
 C/Accession: E75357
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-441 <WHI>
 A/Cross-references: GB:AE002017; GB:AE000513; NID:g6459527; PIDN:AAF11312.1; PID:g645953
 A/Experimental source: strain R1
 C/Genetics:
 A/Gene: DB1756
 A/Map position: 1
 C/Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; try

Query Match 19.6%; Score 332; DB 2; Length 441;
 Best Local Similarity 29.6%; Pred. No. 1.4e-12;
 Matches 107; Conservative 49; Mismatches 110; Indels 96; Gaps 11;

QY 14 QDRFADFPALPLDPSAMVAQVGPVNVNITKLGYNNAVAGAGTGVIVDPNGVLTNNHVA 73
 DB 115 QDQFGF--ALPFDQG-----GGAPATGTGSGFFVNAQGDIVTNNHVD 155

QY 74 GATDINAFVSGGTQGVVDVVDRTQDVAVQLRAGGLPSAAI-----GGGVAVGE 126
 DB 156 GASDITTLHGKQTYKAVIARAPDYLALIR---AEGVPRASIKPLPLGDSKDLVGL 212

QY 127 PVWAMG-----NSGQGGQTPRAVPRGVVAVGQVQASDLSLTGAEETLGLIFDAA 177
 DB 213 KATAMGAPFNLDIFSVEGIISSLERQVP-----VGSREVSQPVQITDAA 256

QY 178 IQPGDSGSPVNGLGQVGVGNNTAASDNFQSQGGQFAIPIGQAMAIAQIRSGGSPV 237
 DB 257 INPNNGSGLLSSAGQVIGVNTQITLGGAGQSGAGVGAIPINTVKRLLPOLQAGK----- 312

QY 238 HIGTAFGLGVVDNN-----GNGARVQVYVVGAPASLGL----- 272
 DB 313 --GVVSPSLGVVFSDDLPPQQLKAAAGLPSGALLQKVVYVPGSPAAGLRGNGKLS 370

Db 371 LPSAQTSISTDGLITAVNGQPLEDAGSLQEAVLATGEGQPLELTVR-RGGKTREVEV 429

QY 323 TL 324

Db 430 TL 431

Search completed: June 22, 2004, 17:24:48
Job time : 10.233 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:46:13 ; Search time 5.49204 Seconds
(without alignments)
3128.737 Million cell updates/sec

Title: US-09-886-349A-4
Perfect score: 1698
Sequence: 1 MHHHHAPPALSDQRFADP.....QTKSGGTRTGNVLAEGPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Swissprot_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	326.5	19.2	458	1	YVTA_BACSU	Q9r911 bacillus su
2	326	19.2	413	1	HTRA_LACHE	Q9z4n7 lactobacill
3	319	18.8	355	1	DEGS_ECOLI	P31137 escherichia
4	315	18.6	408	1	HTRA_LACLA	Q9la06 lactococcus
5	308.5	18.2	513	1	DEGP_BRUME	C8y932 brucella me
6	307.5	18.1	513	1	DEGP_BRUSU	O44597 brucella su
7	305.5	18.0	455	1	DEGO_ECOLI	P26382 escherichia
8	296	17.4	475	1	DEGP_SALTY	Q52894 salmonella
9	296	17.4	504	1	DEGP_RHIME	Q34358 rhizobium m
10	294.5	17.3	449	1	HTRA_BACSU	P09376 bacillus su
11	294.5	17.3	474	1	DEGP_ECOLI	P45129 haemophilus
12	292	17.2	466	1	HTOA_HABIN	P54925 bartonella
13	288.5	17.0	503	1	DEGP_HABHE	P18594 chlamydia t
14	288	17.0	497	1	DEGP_CHLTR	P57322 buchnera ap
15	287	16.9	478	1	DEGP_BUCAI	Q9p197 chlamydia m
16	283.5	16.7	497	1	DEGP_CHLMU	Q9z6t0 chlamydia p
17	282	16.6	488	1	DEGP_CHLPN	O22609 arabidopsis
18	281	16.5	437	1	DEGL_ARATH	O92ja1 rickettsia
19	280	16.5	508	1	DEGP_RICCN	P39668 bacillus su
20	274.5	16.2	400	1	YVXA_BACSU	Q9lu10 arabidopsis
21	272.5	16.0	448	1	DEGS_ARATH	O85291 buchnera ap
22	270.5	15.9	478	1	DEGP_BUCAP	O05942 rickettsia
23	254	15.0	513	1	DEGP_RICPR	O43464 homo sapien
24	248.5	14.6	458	1	HRA2_HUMAN	P44977 haemophilus
25	242	14.3	340	1	DEGS_HABIN	Q92743 homo sapien
26	235.5	13.9	480	1	HEAL_HUMAN	P83110 homo sapien
27	234.5	13.8	453	1	HRA3_HUMAN	Q9j1y5 mus musculu
28	228.5	13.5	458	1	HRA2_MOUSE	P83105 homo sapien
29	225	13.3	476	1	HRA4_HUMAN	Q9d236 mus musculu
30	221.5	13.0	460	1	HRA3_MOUSE	Q9r118 mus musculu
31	221	13.0	480	1	HRA1_MOUSE	Q9sel7 arabidopsis
32	178.5	10.5	321	1	SPPA_ARATH	P55377 rhizobium s
33	175.5	10.3	630	1	Y4BJ_RHISN	

ALIGNMENTS

RESULT 1

ID	YVTA_BACSU	STANDARD;	PRT;	458 AA.
AC	Q9r911; Q35021; Q35039;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Probable serine protease yvta (EC 3.4.21.-).			
GN	YVTA OR YVTB OR BSU33000.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
CO	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=20158875; PubMed=10692364;			
RA	Noone D., Howell A., Devine K.M.;			
RT	"Expression of ykda, encoding a Bacillus subtilis homologue of HtrA,			
RT	is heat shock inducible and negatively autoregulated.";			
RL	J. Bacteriol. 182:1592-1599(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98015415; PubMed=9353931;			
RA	Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;			
RT	"Sequencing of regions downstream of addA (98 degrees) and citG (289			
RL	degrees) in Bacillus subtilis";			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bertoro M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Borriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,			
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,			
RA	Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,			
RA	Guiseppi G., Guy B.J., Haga K., Haech J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,			
RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,			
RA	Kobayashi Y., Koester P., Koningstein G., Krogh S., Kumano M.,			
RA	Kurita K., Lapidis A., Lardinols S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigic C.,			
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,			
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,			
RA	Parro V., Pehl T.M., Portetelle D., Porwollix S., Prescott A.M.,			
RA	Presecan E., Pujic P., Puchelle B., Rapoport G., Rey M., Reynolds S.,			
RA	Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,			
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,			
RA	Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,			
RA	Sorokin A., Taccioni E., Takagi T., Takahashi H., Takenaru K.,			
RA	Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,			
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,			

34	151	8.9	747	1	ELS_BOVIN
35	150.5	8.9	710	1	FIBI_ADEG1
36	150	8.8	457	1	PRTC_STRGR
37	149.5	8.8	515	1	YI40_MYCTU
38	146.5	8.6	864	1	ELS_RAT
39	145.5	8.6	957	1	Y278_MYCTU
40	143	8.4	801	1	Y747_MYCTU
41	141	8.3	2021	1	OMPA_RICCN
42	140	8.2	1150	1	APMU_PIG
43	138	8.1	914	1	WA22_MYCTU
44	137	8.1	778	1	YC3*_MYCTU
45	136	8.0	2249	1	OMPA_RICRI

P04985	bos taurus
O64761	avian adeno
P52320	streptomyce
Q50594	mycobacteri
Q99372	rattus norv
P56877	mycobacteri
O53810	mycobacteri
Q52657	rickettsia
P12021	sus scrofa
O06794	mycobacteri
P71933	mycobacteri
P15921	rickettsia

RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzneger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*,"
RL Nature 390:249-256(1997).
RN [4]
RP TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=20576168; PubMed=11133960;
RA Noone D., Howell A., Collier R., Devine K.M.;
RT "ykda and yvta, Htra-like serine proteases in *Bacillus subtilis*,
RT engage in negative autoregulation and reciprocal cross-regulation of
RT ykda and yvta gene expression,"
RL J. Bacteriol. 183:654-663(2001).
CC -!- FUNCTION: May be involved in processing, maturation, or secretion
CC of extracellular enzymes.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- INDUCTION: Induced by heat shock during exponential growth and by
CC heterologous amylases at the transition phase of the growth cycle.
CC Negatively regulates its own expression.
CC -!- MISCELLANEOUS: Inactivation results in compensating overexpression
CC of htra, especially during stress conditions.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -!- CAUTION: Ref.2 and Ref.3 sequences differ from that shown due to
CC frameshifts in positions 87 and 246 that produce two separate
CC ORFs.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF188296; AA03153.1;
CC EMBL; Z93941; CAB07968.1; ALT_FRAME.
CC EMBL; Z93941; CAB07969.1; ALT_FRAME.
CC EMBL; Z93120; CAB15290.1; ALT_FRAME.
CC Subtilist; BG14155; yvta.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ; 1.
CC PRINTS; PR00089; trypsin; 1.
CC SMART; SM00228; PDZ; 1.
CC PROSITE; PS0106; PDZ; 1.
CC Hydrolyase; Protease; Serine protease; Heat shock; Transmembrane;
CC Complete proteome.
CC DOMAIN 1 71 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 72 92 POTENTIAL.
CC DOMAIN 93 458 EXTRACELLULAR (POTENTIAL).
CC DOMAIN 356 440 PDZ.
CC FT ACT_SITE 187 187 CHARGE RELAY SYSTEM (POTENTIAL).
CC FT ACT_SITE 217 217 CHARGE RELAY SYSTEM (POTENTIAL).
CC FT ACT_SITE 298 298 CHARGE RELAY SYSTEM (POTENTIAL).
CC SEQUENCE 458 AA; 48717 MW; 77551045A865A5CD CRC64;
CC
CC Query Match 19.2%; Score 326.5; DB 1; Length 458;
CC Best Local Similarity 28.4%; Pred. No. 9.1e-12;
CC Matches 101; Conservative 58; Mismatches 140; Indels 57; Gaps 14;
CC
CC 12 LSQDRADFPALPDRSA-WVAQGVQVNNI-NTKLGYNNV-----GAGTGI 57
CC 112 VQSDNFTTAPITNASHADWEDLEPTIVGISNIQTSQNNFTGTGGSSSESSESGTGSV 171
CC
CC 58 VI---DPNGVLLNNHVIAGATINAFSVSGGTGVDVVDRTQDVAVLQRLGAGGLP 114
CC 172 IFKDSKAYIITNNHVEGANKLTV-TLYNETETAKLVGSDTTDLAVLEISGNVKK 230

QY 115 SRAIGGG--VAVGEPTVAMNSGGGGGTTPRAVGRVWALGQTVQASDSLTGAEEITGLI 172
DB 231 VASFGDSSQLRTGKRVIAIGNPLGQOFGTGTGIIISGLNRTIDV-DTQGTVEV--NVL 287
QY 173 QFDAATQPGDSGGPVVNGLGQVVGNTAASDNFQLSQGG---QGFAIPITGQAMATAGAIR 229
DB 288 QTDALINPCNSGGPLINASGVVGIN-----SLKVSSESVESLGFAPISNDVEPIVDQL 342
QY 230 SGGGSPTHVIGTAFILGLGVVNN-----GNGARVQRVVGSPAPASL 271
DB 343 QNGKV-----DRPLGVQMIDMSQVPETYQENTLGLFDQDLGKGVYVKEVQANSFAEKA 396
QY 272 GISTGVDITAVDGAFINSATAMADAL-NGHHPGDDVISVTWQTKSGTGTGNVTIAE 326
DB 397 GIKSEDDIVKNGKDVESADIRQILYKDLKVGDKTTIQ-VLRKGTXTLNATLTK 451

CC RESULT 2
CC HTRA_LACHE
CC ID HTRA_LACHE STANDARD; PRT; 413 AA.
CC AC Q9Z4H7;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Serine protease do-like htra (EC 3.4.21.-).
CC GN HTRA.
CC OS Lactobacillus helveticus.
CC OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
CC OC Lactobacillus.
CC OX NCBI_TaxID=1587;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=53/7;
CC RA MEDLINE=99047559; PubMed=9829922;
CC RX Smeds A., Varmanen P.K., Palva A.M.;
CC RT "Molecular characterization of a stress-inducible gene from
CC Lactobacillus helveticus,"
CC RL J. Bacteriol. 180:6143-6153(1998).
CC CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ005672; CAA06668.1;
CC DR MEROPS; S01.273;
CC DR InterPro; IPR009003; Cys_Ser_trypsin.
CC DR InterPro; IPR001478; PDZ.
CC DR InterPro; IPR001254; Peptidase_S1.
CC DR InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ; 1.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00834; PROTEASES2C.
CC DR SMART; SM00228; PDZ; 1.
CC DR PROSITE; PS0106; PDZ; 1.
CC DR PROSITE; PS0106; PDZ; 1.
CC KW Hydrolyase; Serine protease; Transmembrane.
CC FT TRANSMEM 20 40 POTENTIAL.
CC FT DOMAIN 141 302 CATALYTIC.
CC FT DOMAIN 305 401 PDZ.
CC FT ACT_SITE 150 150 CHARGE RELAY SYSTEM (POTENTIAL).
CC FT ACT_SITE 180 180 CHARGE RELAY SYSTEM (POTENTIAL).
CC FT ACT_SITE 258 258 CHARGE RELAY SYSTEM (POTENTIAL).
CC SEQUENCE 413 AA; 42647 MW; B16B677991C88707 CRC64;

CC Query Match 19.2%; Score 326; DB 1; Length 413;
CC Best Local Similarity 31.6%; Pred. No. 8.7e-12;

```

Matches 90; Conservative 50; Mismatches 121; Indels 24; Gaps 8;
QY 42 NTKLGNNAVAGTGIIV--DPNGVLTNNHVIAGATDINAFSGVGQTGVDDVVGYDR 98
DB 119 SSXNGKLEYSSEGGVYVYKSGKGIYVNNHVISGSDAVQVL-LANGKTVNAKVVGKDS 177
QY 99 TODVAVLQIRAGGLPSAIGGG--VAVGEPVVMNGSGGGGTTPRVPGRVVALGQTVQ 156
DB 178 TTDLAVLSDAKYVTTQAGFGSKHLEAGQTVIAGSPILGSEYASTVTQGIISAPARTIS 237
QY 157 ASDSLTGARETLNGLQFDAAITOPGSGGPPVNGLVGVGMNT--AASDNFQLSGGQGF 214
DB 238 TS---SGNQOT---VLTQDAINPNSGALVNSAGQVIGNSMKLAQSSDTSVEGNAP 291
QY 215 AIPIGQMAIAGQIRSGG--GSFTVHIGTAPLIG-----LGVVDNNGNGARVQVRVG 264
DB 292 AIPNSNEWTVNELVKKGKTRPOLGVRVIALQGIPEGVRSRLKIKSNLKNGIYIAFVSR 351
QY 265 SAPAASLGTSGDVTAVGAPINSATAMADALNGHHPGDVISVT 309
DB 352 NGSAAAGIKSGDVTIKVDPKKVEDVASLHSLYSHKVGDTVNVT 396

RESULT 3
DEGS_ECOLI STANDARD; PRT; 355 AA.
AC F31137;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protease degs precursor (EC 3.4.21.-).
GN DEGS OR HHOB OR HTHF OR B3235 OR Z4594 OR ECS4108.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Bass S., Gu Q., Goddard A.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=96165272; PubMed=8576051;
RA Waller P.R., Sauer R.T.;
RT "Characterization of degQ and degS, Escherichia coli genes encoding
RT homologs of the DegP protease.";
RL J. Bacteriol. 178:1146-1153(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1245-1247(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]

```

```

SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kunara S., Shiba I., Hattori M., Shinagawa H., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP PRELIMINARY SEQUENCE OF 82-355 FROM N.A.
RX MEDLINE=88105815; PubMed=3222223;
RA Vogel R.F., Entian K.-D., Mecke D.;
RT "Cloning and sequence of the mdh structural gene of Escherichia coli
RT coding for malate dehydrogenase.";
RL Arch. Microbiol. 149:36-42(1987).
RN [7]
RP IDENTIFICATION.
RA Bazan J.F., Fletterick R.J.;
RT "Structural and catalytic models of trypsin-like viral proteases.";
RL Semin. Virol. 1:311-322(1990).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2c.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
EMBL; U15661; AAC43993.1; -
EMBL; U32495; AAC4006.1; -
EMBL; U18997; AAA58037.1; -
EMBL; AE000402; AAC76267.1; -
EMBL; AE005551; AAG58363.1; -
EMBL; AP002564; BAB37531.1; -
EMBL; M24777; -; NOT_ANNOTATED_CDS.
DR PIR; D91142; D91142.
DR PIR; JC6052; JC6052.
DR MEROPS; S01.275; -
DR EcoGene; EGI1652; hhoB.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ_Ser.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1c.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolyase; Serine protease; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 28
FT CHAIN 29 355
FT DOMAIN 281 326
FT ACT_SITE 96 96
FT ACT_SITE 126 126
FT ACT_SITE 201 201
FT CONFLICT 253 253
FT CONFLICT 307 307
SQ SEQUENCE 355 AA; 37581 MW; D091B4D65E8FE1CC CRC64;
Query Match 18.8%; Score 319; DB 1; Length 355;
Best Local Similarity 31.2%; Pred. No. 1.9e-11;
Matches 109; Conservative 43; Mismatches 137; Indels 50; Gaps 13;
QY 12 LSQRFADFPALPDPSAMVAGVQVYVNTKGYNNVAGV-----GTCVID 60
DB 32 LSTPQDGTDETPASYNLAVERAAPVVV-----YNRGLNTNSHNQLRTLGSGVMD 86

```

QY 61 PNGVLNNHVIAGATDINAFSVGGQTYGVVVGVDRTQDVAVLQRLGAGGLPSAAIGG 120
Db 87 QRGVITIKHVIINDAQI-IVALQDGRVFEALLVGSLSLTLAVLKINATGGLPTIPINA 145
QY 121 GVA--VSEPVMANGSGGGTTPRAVGRVVALGQTV-QASDSLTAETLN-----GLI 172
Db 146 RRVPHIGDVLAINP-----YNLQQTITQIGIISATG-RIGLNPTRGRNFL 190
QY 173 QFDAAIQPGSGGVVNGLGQVGMNTAASD--NFOLSGGGQGFAPF-----IG 219
Db 191 QTDASINHGSGGALVNSLSELGMLNTLSPKNSDGTPEGIGFAPFPQATKIMDKLIR 250
QY 220 QAMAIQIRSGGSPVHIGTFAFLGLGVVDNNGNGARVQRVVGSAAPASLIGISTGDVI 279
Db 251 DGRVIRYIGIGGR-----EIAPLHAQGGI--DQLQGIIVNEVSPDGPAAAGIQVNDLI 304
QY 280 TAVDGPAINSATAMADALNGHHPGDVISTWQKSGGTRTGNVTLABGP 328
Db 305 ISVDNKAISALETMDQVAIRPGSVIPV-VNRDDKQLTLQVTIQBYP 352

RESULT 4
HTRA_LACLA
ID HTRA_LACLA STANDARD; PRT; 408 AA.
AC Q9LA06;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine protease do-like htra (EC 3.4.21.-) (HTRALL).
GN HTRA OR L2136.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN
SEQUENCE FROM N.A., AND CHARACTERIZATION.
RP STRAIN=IL1403;
RX MEDLINE=20177820; PubMed=10712686;
RA Poquet I., Saint V., Seznec E., Simoes N., Bolotin A., Gruss A.;
RT "Htra is the unique surface housekeeping protease in Lactococcus
RT lactis and is required for natural protein processing."
RL Mol. Microbiol. 35:1042-1051(2000).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Winkler P., Mauger S., Jaillon O., Malarme K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
CC -!- FUNCTION: Degrades abnormal exported proteins. Needed for the pro-
CC PEPTIDE PROCESSING OF A NATURAL PRO-PROTEIN AND FOR MATURATION OF
CC A NATIVE PROTEIN. RESPONSIBLE FOR THE HOUSEKEEPING OF EXPORTED
CC PROTEINS.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; AF155705; AAF61294.1; -
CC EMBL; AE006442; AAK06234.1; -
CC PIR; H86891; H86891.
CC MEROPS; S01.273; -
CC InterPro; IPR009003; Cys_ser_trypsin.
CC InterPro; IPR001478; PDZ-
CC InterPro; IPR001254; Peptidase_S1.

DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydroxylase; Serine protease; Transmembrane; Complete proteome.
FT TRANSMEM 6 26
FT DOMAIN 88 284
FT ACT_SITE 302 383
FT ACT_SITE 127 127
FT ACT_SITE 157 157
FT ACT_SITE 239 239
SQ SEQUENCE 408 AA; 41648 MW; 581B90B55A7DF851 CRC64;
Query Match 18.6%; Score 315; DB 1; Length 408;
Best Local Similarity 31.6%; Pred. No. 3.6e-11;
Matches 94; Conservative 51; Mismatches 118; Indels 34; Gaps 11;
QY 54 GTGIVIDPNG---VVLNNHVIAGATDINAFSVGGQTYGVVVGVDRTQDVAVLQRLG- 109
Db 108 GSGVYKKSGGDYVVTNYHYIAGNSSLDVLLSG-GQVKASVVGDEYTDLAVLKISSE 166
QY 110 -AGGLPSAAIGGVAVGPPVAVMGNSGGQGTTPRAVGRVVALGQTVQASDSLTAETL 168
Db 167 HVKDVATFADSSKLTIGEPALVGSPLGSQPANTATGILSATSRQVTLTQE-NGQTNI 225
QY 169 NGLIQFDAAIQPGSGGVVNGLGQVGMNTA---ASDNFOLSGGGQGFAPFPGQAMVIA 225
Db 226 NA-IQTDAAINPGNSGGALINIEGVIGITQSKITTEDGSTVEGLGFAIPSDVNNII 284
QY 226 QGIRSGGSPVHIGTFAFLGLGVVD-----NNGN-----GARQVRVGSAPAA 269
Db 285 NKLEADG-----KISRPA-LGIRWVDLSQLSTNDSQKLPSSVTGGVVVSVSGSLPAA 338
QY 270 SLIGISTGDVITAVDGPAINSATAMADALNGHHPGDVISTWQKSGGTRTGNVTLAE 326
Db 339 SAGLKAGDVITKVGDTAVTSSTDLSALYSHNINDTVKVTY-RDGKSNATADVLSK 394
RESULT 5
DEGP BRUME
ID DEGP BRUME STANDARD; PRT; 513 AA.
AC Q8VG32;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
DEGP OR HTRA OR BMEI1330.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapatal V., Redkar R.J., Patra G., Mujar C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyrides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

[illegible]

RESULT 9

FT CONFLICT 464 504 KTVTSPTTSSPART (IN REF. 1).
 FT BEST LOCAL SIMILARITY 37.4%; Score 296; DB 1; Length 504;
 SQ SEQUENCE 504 AA; 53035 MW; D7E82B9981E23C CRC64;
 Matches 98; Conservative 39; Mismatches 116; Indels 38; Gaps 12;
 Query Match
 54 GTGIVIDNGVLLNNHVIAGATDINAFSV--GGQTYGVGVVGVYDRTQDAVLQL--RG 109
 DB 124 GSGFFITDGYLVNTHV---SDGSAFTVIMNDGTDLDAKVGKDSRTDLAVLKVDKDR 180
 QY 110 AGGLPSAIGGVAVGVVAVMGNSGGGTFPRVAVGRVVALGQTVASLSLTCAEETLN 169
 DB 181 KFTVVSFADDEKVRVGVVAVGVPFGLGGVTVA--GLISARGRDIGSG-----PYD 230
 QY 170 GLIQDAAIQDGGGVVNGVGLGVWGMNTAASDNFQSQGGG--FAIFIGQAMAIAGQ 227
 DB 231 DYLQVDAVNRGNSGGPTFNLSGEWGINTAL---FSPSGGVGIAFAIPASVAKDVWDS 287
 QY 228 IRSGG---GSPVTHIGFTA-----FLGLGVWNGNGARVQRVVSAPASLIGISTGV 278
 DB 288 LIKDGTVSRGMLGVQIQFVTKDIAESLGL-----SEANGALVWEPQAGSGEKAGIKNGDV 343
 QY 279 ITAVDGAIPINSATAMADALNHHPGDVSVT-WOTKSGGTRTGNVTLAEGP 328
 DB 344 VTALNGEPVDPRLARRVAAALRGSTAEVTLW--RSKSETVNLGIGTLP 392
 RESULT 10
 HTRA_BACSU STANDARD; PRT; 449 AA.
 AC 034358;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 10-OCT-2003 (Rel. 42, Last annotation update)
 GN HTRA OR BSUI2900.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Devine K.M.;
 RL "Sequence of the Bacillus subtilis genome between xlyA and ykoR.";
 RN Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Etian K.D., Errington J., Fabret C., Ferracioli S., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Funa S., Gallizzi A., Galleron N.,
 RA Gim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Juris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigic C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Neone D., O'Reilly M., Ozawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis.";
 RL Nature 390:249-256(1997).
 RN [3]
 RP TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX MEDLINE=20158875; PubMed=10692364;
 RA Noone D., Howell A., Devine K.M.;
 RT "Expression of ykdA, encoding a Bacillus subtilis homologue of HtrA,
 RT is heat shock inducible and negatively autoregulated.";
 RL J. Bacteriol. 182:1592-1599(2000).
 RN [4]
 RP TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX MEDLINE=20576168; PubMed=11133960;
 RA Noone D., Howell A., Collier R., Devine K.M.;
 RT "ykdA and yvtA, HtrA-like serine proteases in Bacillus subtilis,
 RT engage in negative autoregulation and reciprocal cross-regulation of
 RT ykdA and yvtA gene expression.";
 RL J. Bacteriol. 183:654-663(2001).
 RN [5]
 RP TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX MEDLINE=21439741; PubMed=1155295;
 RA Hytynlaenen H.-L., Bolhuis A., Darmon E., Muukkonen L., Koski P.,
 RA Vitikainen M., Sarvas M., Pragai Z., Bron S., van Dijk J.M.,
 RA Kontinen V.P.;
 RT "A novel two-component regulatory system in Bacillus subtilis for the
 RT survival of severe secretion stress.";
 RL Mol. Microbiol. 41:1159-1172(2001).
 CC -!- FUNCTION: May be involved in processing, maturation, or secretion
 CC of extracellular enzymes.
 CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
 CC -!- INDUCTION: Transcription is cssS dependent. Induced by heat shock
 CC during exponential growth and by heterologous amyases at the
 CC transition phase of the growth cycle. Negatively regulates its own
 CC expression during exponential growth and during heat shock.
 CC -!- MISCELLANEOUS: Inactivation results in compensating overexpression
 CC of yvtA, especially during stress conditions.
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 1 PDZ/CHR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AJ002571; CAA05570.1; -;
 DR EMBL; Z99110; CAB13147.1; -;
 DR PIR; A69643; A69643.
 DR MEROPS; S01-273; -;
 DR SubtilList; B012608; htra.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00334; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 DR HydroLase; Protease; Serine protease; Heat shock; Transmembrane;
 KW Complete proteome.

FT	DOMAIN	1	44	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	45	67	POTENTIAL.
FT	DOMAIN	68	439	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	103	108	POLY-SER.
FT	DOMAIN	146	152	POLY-SER.
FT	DOMAIN	348	437	PDZ.
FT	ACT SITE	179	179	CHARGE RELAY SYSTEM (POTENTIAL).
FT	ACT SITE	209	209	CHARGE RELAY SYSTEM (POTENTIAL).
FT	ACT SITE	290	290	CHARGE RELAY SYSTEM (POTENTIAL).
SQ	SEQUENCE	449 AA;	47713 MW;	E12B07A9018EB414 CRC64;
Query Match				
Best Local Similarity 29.3%; Pred. NO. 5.6e-10;				
Matches 93; Conservative 52; Mismatches 127; Indels 45; Gaps 11;				
QY	28	SANVAGVQGVNINIKLGNNA-----	VGAGTGIVI-DPNG-	-VVLINN 69
DB	119	SDMVEDLSAIVGICITMLQAGSSLSFGSSSSDSTESGSGVGIFKENGKAVIINN		178
QY	70	HVIAGATDINAFSVGSGQTYGVVDVGYDRTQDAVLQLRAGGLPSAIG--GGVAVGEP		127
DB	179	HVVEGASSLKV-SLYDGTETAKLVGSDSLTDLAVLIQISDDHVTKVANFGDSSDLRTGET		237
QY	128	VVAMGNSGGGTPRAVGRVVALGQVQASDSLTCAEETLNGLIQFDAAIOPGSGGPV		187
DB	238	VTAIGDPLGKDSRTVTQGVSGVDETVMSS---TSAGETSNVITQDAAINPGSGGGL		294
QY	188	VNLGIGQWGNNT--AASDNFQLSQGGQGAIPICQAMAIAGTRSGGSGFTVHIGTAPL		245
DB	295	LNTDGIKIVGNSKISDEDDVE-----GIGFAIPNDVKPIABELLSKGQIERPIVG-VSML		349
QY	246	GLGVVDNN-----	NGARVQRVVGSAAPASIGSTGVDTAVDQAPINSATA	292
DB	350	DLEQVPQNYQEGTGLFGSLNKGVIYREVASGSPAERKAGLKAEDIIIGLKGEIDTSGE		409
QY	293	MADAL-NGHHPGDVISV	308	
DB	410	LRNLKYDAKIGDTVEV	426	

RESULT 11

ID	DEGP_ECOLI	STANDARD;	PRT;	474 AA.
AC	P09376; E15724;			
DT	01-MAR-1989 (Rel. 10, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Protease do precursor (EC 3.4.21.-)			
GN	DEGP OR HTRA OR PTD OR B0161 OR Z0173 OR ECS0165.			
OS	Escherichia coli, and			
OS	Escherichia coli O157:H7.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562, 83334;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / MG1655;			
RC	MEDLINE=97426617; PubMed=9278503;			
RC	STRAIN=K12 / W3110;			
RC	MEDLINE=94261430; PubMed=8202364;			
RC	Fujita N., Mori H., Yura T., Ishihama A.;			
RA	"Systematic sequencing of the Escherichia coli genome: analysis of			
RT	the 2.4-4.1 min (110,917-193,643 bp) region.;"			
RT	coli: a sigma 32-independent mechanism of heat-inducible			
RT	transcription.;"			
RL	Nucleic Acids Res. 16:10053-10067 (1988).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / W3110;			
RC	MEDLINE=94261430; PubMed=8202364;			
RC	Fujita N., Mori H., Yura T., Ishihama A.;			
RA	"Sequence analysis and regulation of the htrA gene of Escherichia			
RT	coli: a sigma 32-independent mechanism of heat-inducible			
RT	transcription.;"			
RL	Nucleic Acids Res. 22:1637-1639 (1994).			
RN	[3]			
RP	SEQUENCE FROM N.A.			

RC	STRAIN=K12 / MG1655;			
RC	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Ferna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RA	Mau B., Shao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12.;"			
RL	Science 277:1453-1474 (1997).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,			
RA	Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,			
RA	Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,			
RA	Davis R.W.;			
RL	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=O157:H7 / EDL933 / ATCC 700927;			
RC	MEDLINE=21074935; PubMed=11206551;			
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,			
RA	Rose D.J., Mayhew G.P., Evans P.S., Gregor J., Kirkpatrick H.A.,			
RA	Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,			
RA	Grobbeck B.J., Davis N.W., Lim A., Dimalanta E.T., Potamocous K.,			
RA	Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,			
RA	Weich R.A., Blattner F.R.;			
RT	"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.;"			
RL	Nature 409:529-533 (2001).			
RN	[6]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=O157:H7 / RIMD 0509952;			
RC	MEDLINE=21156231; PubMed=11258796;			
RA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,			
RA	Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,			
RA	Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,			
RA	Kuhara S., Shiba T., Hattori M., Shinagawa H.;			
RT	"Complete genome sequence of enterohaemorrhagic Escherichia coli			
RT	O157:H7 and genomic comparison with a laboratory strain K-12.;"			
RL	DNA Res. 8:11-22 (2001).			
RN	[7]			
RP	SEQUENCE OF 1-50 FROM N.A.			
RC	STRAIN=K12;			
RC	MEDLINE=90323597; PubMed=2165018;			
RA	Quirk S., Bhatnagar S.K., Bessman M.J.;			
RT	"Primary structure of the deoxyguanosine triphosphate			
RT	triphosphohydrolase-encoding gene (dgt) of Escherichia coli.;"			
RL	Gene 89:13-18 (1990).			
RN	[8]			
RP	SEQUENCE OF 1-16 FROM N.A.			
RC	MEDLINE=90207273; PubMed=2157212;			
RA	Wurgler S.M., Richardson C.C.;			
RT	"Structure and regulation of the gene for dGTP triphosphohydrolase			
RT	from Escherichia coli.;"			
RL	Proc. Natl. Acad. Sci. U.S.A. 87:2740-2744 (1990).			
RN	[9]			
RP	SEQUENCE OF 27-39, AND CHARACTERIZATION.			
RC	MEDLINE=90202693; PubMed=2180903;			
RA	Lipinska B., Zylcz M., Georgopoulos C.;			
RT	"The HtrA (DegP) protein, essential for Escherichia coli survival at			
RT	high temperatures, is an endopeptidase.;"			
RL	J. Bacteriol. 172:1791-1797 (1990).			
RN	[10]			
RP	SEQUENCE OF 27-30.			
RC	STRAIN=K12 / W3110;			
RC	MEDLINE=98263247; PubMed=9600841;			
RA	Wilkins M.R., Gasteiger E., Tonelli L., Ou K., Tyler M.,			
RA	Sanchez J.-C., Gooley A.A., Walsh B.J., Bairoch A., Appel R.D.,			
RA	Williams K.L., Hochstrasser D.F.;			
RT	"Protein identification with N and C-terminal sequence tags in			
RT	proteome projects.;"			
RL	J. Mol. Biol. 278:599-608 (1998).			
RN	[11]			
RP	IDENTITY OF HTRA AND PROTEASE DO.			
RP	MEDLINE=9122240; PubMed=2025286;			

Seol J.H., Woo S.K., Jung E.M., Yoo S.J., Lee C.S., Kim K.J.,
Tanaka K., Ichihara A., Ha D.B., Chung C.H.,
"Protease do is essential for survival of *Escherichia coli* at high
temperatures: its identity with the *htrA* gene product.",
RL Biochem. Biophys. Res. Commun. 176:730-736(1991).
RP DISULFIDE BOND.
RX MEDLINE=22760311; PubMed=12878036;
RA Skorko-Glonek J., Zurawa D., Tanfani P., Seire A., Wawrzynow A.,
RA Narkiewicz J., Bertoli E., Lipinska B.,
RT "The N-terminal region of *HtrA* heat shock protease from *Escherichia*
RT coli is essential for stabilization of *HtrA* primary structure and
RT maintaining of its oligomeric structure.",
RL Biochim. Biophys. Acta 1649:171-182(2003).
CC -!- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IT CAN DEGRADE
CC ICIA, ADA, CASEIN AND GLOBIN. SHARED SPECIFICITY WITH DEGO.
CC -!- SUBUNIT: Multimeric.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- INDUCTION: By heat shock.
CC -!- MISCELLANEOUS: *HtrA* IS INDISPENSABLE FOR BACTERIAL SURVIVAL AT
CC TEMPERATURES ABOVE 42 DEGREES CELSIUS.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M36536; AAA23994.1; -;
CC EMBL; X12457; CAA30997.1; -;
CC EMBL; D26562; BAB98738.1; -;
CC EMBL; AE000125; AAC73272.1; -;
CC EMBL; U70214; AAB08591.1; -;
CC EMBL; AE005192; AAG54465.1; -;
CC EMBL; AP002550; BAB33588.1; -;
CC EMBL; M29955; AAA23717.1; -;
CC EMBL; M31772; AAA23680.1; -;
CC PIR; B85500; E85500.
CC PIR; E90849; E90649.
CC PIR; S45229; S45229.
CC PDB; 1KY9; 03-APR-02.
CC MEROPS; S01.273; -;
CC SWISS-2DPAGE; P09376; COLI.
CC EcoGene; EG10463; degP.
CC InterPro; IPR009003; Cys_ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ; 2.
CC Pfam; PF00089; tryptsin; 1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SM00228; PDZ; 2.
CC PROSITE; PS0106; PDZ; 2.
CC Hydrolase; Serine protease; Heat shock; Periplasmic; Repeat; signal;
KW Complete proteome; 3D-structure.
FT SIGNAL 1 26
FT CHAIN 27 474
FT DOMAIN 280 371
FT PDZ 1.
FT PDZ 2.
FT ACT_SITE 131 131
FT ACT_SITE 161 161
FT ACT_SITE 236 236
FT ACT_SITE 83 95
FT DISULFID 10 10
FT CONFLICT 46 46
FT CONFLICT 10 10
FT CONFLICT 192 192
FT CONFLICT 467 467
FT CONFLICT 192 192
FT CONFLICT 467 467
FT REF. 1).
FT STYLMQ -> RHLFVNAVSLNPLKTRGSPNVL (IN
FT REF. 1).

SQ SEQUENCE 474 AA; 49354 MW; 5482E596F74B6D5F CRC64;
Query Match 17.3%; Score 294.5; DB 1; Length 474;
Best Local Similarity 33.6%; Pred. No. 5.9e-10;
Matches 89; Conservative 42; Mismatches 81; Indels 53; Gaps 11;
QY 54 CTGIVIDPN-GVLTNNHVIACATINAFSVSGGTGYGVVGVYDRTQDVAVLQRGAGG 112
Db 114 GSGVIIDAKGVYVNNHVDNATVIKV-QLSDGRKFDKMKVGDPRSDIALIQNPKN 172
QY 113 LPS--AAIGGVAVGEPVVMGNSGGQGTTPRAVPRVVALCQTVQASDSLTGAB-ETLN 169
Db 173 LTAIKWADSDALRVGDYTVAGNPFGLGET--VTSGIVSALGR-----SGLNAENYE 222
QY 170 GLIQDAAIQPDGSGPVVNGLGQVVGNTA--ASDNFQLSQGGGPAIP-----IQQA 221
Db 223 NFQIQDAAINRGNSGGALVNLNGELIGINTAILAPDGGNI---GIGFAIPSNMVKNTSQ 279
QY 222 MAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGN-----ELGIMGTLSLAKAMKVDQAQGFVSQVLPNSSA 268
Db 280 MVEYGVQVREG-----
QY 269 ASLGISTGVITAVDQAPINSATAM 293
Db 326 AKAGIKAGDVITSLNGKPISSPAAL 350
RESULT 12
HTOA_HAEIN
ID HTOA_HAEIN STANDARD; PRT; 466 AA.
AC P45129;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable periplasmic serine protease do/hhoA-like precursor
DE (EC 3.4.21.-).
GN H11259.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]_TaxID=727;
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Gloeck A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fire L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of *Haemophilus influenzae*
RT Rd.",
RL Science 269:496-512(1995).
CC -!- SUBCELLULAR LOCATION: Periplasmic (potential).
CC -!- SIMILARITY: SEEMS TO BE A INTERMEDIATE FORMS BETWEEN E.COLI HTRA
CC (PROTEASE DO) AND HHOA.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U32805; AAC22906.1; -;
CC PIR; A64113; A64113.

DE immunogenic protein) (SK59).
GN DEGP OR HTRA OR CT823.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1] _SEQUENCE FROM N.A.
RP STRAIN=Setovar L2.
RC MEDLINE=90337348; PubMed=2379836;
RX Kahane S., Weinstein Y., Sarov I.,
RA "Cloning, characterization and sequence of a novel 59-kDa protein of
RT Chlamydia trachomatis";
RL Chlamydia trachomatis";
RN Gene 90:61-67 (1990).
RN [2]
RP _SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/Cx;
RX MEDLINE=9900809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis";
RL Science 282:754-759 (1998).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DRH domains.
CC -!- CAUTION: REF.1 SEQUENCE WAS INCORRECT: DUE TO SEQUENCING ERRORS
CC THE AUTHORS TRANSLATED THEIR PUTATIVE 59 kDa IMMUNOGENIC PROTEIN
CC ON THE COMPLEMENTARY STRAND TO THAT OF WHAT SEEMS TO BE THE REAL
CC PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M31119; AAA23116.1;
DR EMBL; M31119; AAA23116.1;
DR PIR; H71465; H71465.
DR PHCI-2DPAGE; P1584;
DR InterPro; IPR003003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR008256; Peptidase_S1B_V8.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SMO0228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
DR Hydroxylase; Serine protease; Repeat; Signal; Antigen;
KW Complete proteome.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 497 PROBABLE SERINE PROTEASE DO-LIKE.
FT DOMAIN 128 289 CATALYTIC.
FT DOMAIN 290 381 PDZ 1.
FT DOMAIN 394 485 PDZ 2.
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 497 AA; 53244 MW; 86A5E31B84A36BA CRC64;
Query Match 17.0%; Score 288; DB 1; Length 497;
Best Local Similarity 32.5%; Pred. No. 1.4e-09;
Matches 95; Conservative 45; Mismatches 112; Indels 40; Gaps 11;
QY 54 GTGIVDPNGVLTNNHVIAGATDINAFVSGSGQYGVVDVYDQVAVQLRGAGGL 113
DB 127 GTGIVSESDGYVTHNHVVEDAGKHV-TLHDGQKYTAKIVGLDKPTDLAVIKIQ-AEKL 184

QY 114 PSAAIGGG--VAVGEPVAMGNSGGGGTTPRAVPGRVVALGQT-----VOASDSLTAET 167
DB 185 PFLTFGNSDQLQIGDMAIAIGNPFGIQAAT--VTVGVISAKGRNQLHIVDFD----- 234
QY 168 LNLGIQFDAAIOPGDSGGPVNGLGOVVMNTAASDNFOLSQ--GGGFAIPIGAMAIA 225
DB 235 ---FIOTDAIINPGSGGELLNNGVIGVNTAIVSG---SGYIGIGFAISLMKRV 288
QY 226 GQIRSGGSGPTVHIGPTAFGLGLGVVDN-----NGNGARVQRVVGSAASLIGISG 276
DB 289 DQLISDQVTRGFLGVT-----LQPIDSELATCYKLEKVVYGVALTVDVKGSPAAGLQGE 344
QY 277 DVITAVDGAIPINSATAMADALNGHPGDVISVTWQTKSGGTGTGNTVLAEP 328
DB 345 DVIVAYNGKEVESLSALRNAISLMMEGTRV-VLKIVREGKTIPIPTVTQIP 395

RESULT 15
ID DEGP_BUCAI STANDARD; PRT; 478 AA.
AC P57322;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR BU228
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1] _SEQUENCE FROM N.A.
RP STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS";
RL Nature 407:81-86 (2000).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DRH domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP001118; BAB12943.1;
DR MEROPS; S01.273;
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SMO0228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 1.
DR Hydroxylase; Serine protease; Repeat; Signal; Complete proteome.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 478 PROBABLE SERINE PROTEASE DO-LIKE.
FT DOMAIN 116 254 CATALYTIC.
FT DOMAIN 281 372 PDZ 1.
FT DOMAIN 387 469 PDZ 2.
FT ACT_SITE 133 133 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 163 163 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (POTENTIAL).
FT DISULFID 87 99 BY SIMILARITY.
SQ SEQUENCE 478 AA; 52230 MW; 868E8732CAC50629 CRC64;

Query Match 16.9%; Score 287; DB 1; Length 478;
Best Local Similarity 30.9%; Pred. No. 1.6e-09;
Matches 89; Conservative 50; Mismatches 109; Indels 40; Gaps 12;

QY	54	GTGIVDPN-GVVLNNHVIAGATDINAFSGVGQTYGVVDYDRTQDVAVLQLRGAGG	112
Db	116	GGVLIINADKGYAVTNHVNENANKIQV-QLSDGRYEAREVIGKDSRSDIALIQLKNANN	174
QY	113	LPSAAI--GGGVAVGEPVVMGNSGGGTPRAVPRVVALGQTVQASDSLTGAE-ETLN	169
Db	175	LSEIKIADSDNLRVGDYTVAINPYGLGET--VTSGIISALGR-----SGLNIEHYE	224
QY	170	GLIOPDAATQPGDSGPPVNG-GQVVGMNTA--ASDNFOLSGOGGPAIPICQAMAIAGQ	227
Db	225	NFIQTDAALNRGNSGALVNLKGLIGINTAILAPDGGNI---GIGPAIFCNWVKNLTAQ	281
QY	228	IRSGGGSPTVHIGPTAFLGLGVVD-----NNGGARVQVRVWGSAPAAISLGISTGDVI	279
Db	282	MQVFG--QVRRGELGIMGMELNSDLAQIMKINSQKGFVSRVLPNSSAFEAGIKAGDII	338
QY	280	TAVDGAPINSATAMADALNGHHFGDVISVTWTKSGGTGTG---NVTL	324
Db	339	ISLNRKPISSFSLSRAEIGS-----LPVATKMWELGVFREGRIKNTIV	380

Search completed: June 22, 2004, 17:16:47
Job time : 6.58295 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:55:54 ; Search time 26.9826 Seconds
(without alignments)
3858.816 Million cell updates/sec

Title: US-09-886-349A-4
Perfect score: 1698
Sequence: 1 MHFHHHAPPALSQDRFADF.....QTKSGGTRGTGNTVLAEGPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 31518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL 25:*
- 1: sp_archea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_virus:*
 - 16: sp_bacteriap:*
 - 17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1645	96.9	355	16 Q07175	O07175 mycobacteri
2	1645	96.9	355	16 Q7U289	Q7U289 mycobacteri
3	1231.5	72.5	361	2 Q50320	Q50320 mycobacteri
4	1198	70.6	354	16 Q9CCY9	Q9CCY9 mycobacteri
5	460.5	27.1	446	16 Q8VKA4	Q8VKA4 mycobacteri
6	460.5	27.1	464	16 Q7U0X2	Q7U0X2 mycobacteri
7	459.5	27.1	464	16 Q53896	Q53896 mycobacteri
8	450	26.5	382	16 Q9CD67	Q9CD67 mycobacteri
9	450	26.5	452	2 Q2Z566	Q2Z566 mycobacteri
10	397.5	23.4	519	16 Q93J30	Q93J30 streptomyc
11	396	23.3	375	16 Q8DG87	Q8DG87 synchococc
12	393.5	23.2	542	16 Q9FBK9	Q9FBK9 streptomyc
13	384	22.6	472	16 Q82FM9	Q82FM9 streptomyc
14	384	22.6	619	16 Q82I18	Q82I18 streptomyc
15	379	22.3	473	16 Q8F817	Q8F817 corynebacte
16	374	22.0	394	16 P72780	P72780 synchocyst

17	368	21.7	407	16 Q8YTF9	Q8YTF9 anabaena sp
18	362	21.3	326	16 Q8DMV9	Q8DMV9 synchococc
19	357.5	21.1	675	16 Q8G6T3	Q8G6T3 bifidobacte
20	357	21.0	413	16 Q8NS10	Q8NS10 corynebacte
21	354.5	20.9	395	16 Q7V5C8	Q7V5C8 prechloroco
22	353.5	20.8	391	16 Q89915	Q89915 clostridium
23	353	20.8	433	16 Q37GB5	Q37GB5 clostridium
24	353	20.8	447	16 Q8R756	Q8R756 thermococci
25	352	20.7	452	16 P73354	P73354 synchocyst
26	351	20.7	339	16 Q89RP2	Q89RP2 bradyrhizob
27	350	20.6	525	16 Q8PMV4	Q8PMV4 xanthomonas
28	349	20.6	500	2 Q9KJN6	Q9KJN6 myxococcus
29	348	20.5	374	16 Q7U435	Q7U435 synchococc
30	348	20.5	490	16 Q8XP75	Q8XP75 ralsstonia s
31	346.5	20.4	389	16 Q8DL28	Q8DL28 synchococc
32	346	20.4	371	16 Q31388	Q31388 bradyrhizob
33	345.5	20.3	525	16 Q89QJ8	Q89QJ8 bradyrhizob
34	343.5	20.2	432	16 Q7U8K9	Q7U8K9 synchococc
35	339	20.0	514	16 Q87C10	Q87C10 xylella fas
36	338	19.9	514	16 Q9PBA3	Q9PBA3 xylella fas
37	336.5	19.8	377	16 Q7U604	Q7U604 synchococc
38	335.5	19.8	362	16 Q9S2K5	Q9S2K5 streptomyc
39	335.5	19.8	465	16 Q92QB6	Q92QB6 rhizobium m
40	335	19.7	511	2 Q8RTK2	Q8RTK2 xanthomonas
41	335	19.7	525	16 Q8PB56	Q8PB56 xanthomonas
42	333	19.6	416	16 Q8YVH0	Q8YVH0 anabaena sp
43	332	19.6	398	16 Q83752	Q83752 treponema p
44	332	19.6	441	16 Q9TK4	Q9TK4 deinococcus
45	331	19.5	371	16 Q7VA24	Q7VA24 prochlorococ

ALIGNMENTS

RESULT 1

O07175 PRELIMINARY; PRT; 355 AA.

ID O07175; (TrEMBLrel. 04, Created)

AC O07175; (TrEMBLrel. 04, Last sequence update)

DT 01-JUL-1997

DE Hypothetical protein (Serine protease, putative).

GN PEPA OR RV0125 OR MTC1418B.07 OR MT0133.

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

CC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H37RV;

RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,

RA Hornaby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,

RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;

RT "Deciphering the biology of Mycobacterium tuberculosis from the

RT complete genome sequence."

RN [2]

RP Nature 393:537-544 (1998).

RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

RA Peterson J.F., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Ormrod M.D., Salzberg S.L.,

RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,

RA Bishai W.;

RT "Whole genome comparison of Mycobacterium tuberculosis clinical and

RT laboratory strains."

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

```

CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; Z96071; CAB09453.1; -.
DR EMBL; AE006925; AAK44357.1; -.
DR PIR; F70983; F70983.
DR TIGR; MT0133; -.
DR Tuberculin; RV0125; -.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR008256; Peptidase_S1b_v8.
DR InterPro; IPR001940; Peptidase_S1c.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00106; PDZ; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Hypothetical protein; Serine protease; Protease;
KW Complete proteome.
SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 96.9%; Score 1645; DB 16; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.4e-81;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFALPLDPSAMVAQGVQVNNITKLGYNNAVAGAGTIVDPNGVVL 67
DB 33 APPALSQDRFADFALPLDPSAMVAQGVQVNNITKLGYNNAVAGAGTIVDPNGVVL 92
QY 68 NNHVIAGATDINAFSVSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
QY 128 VVMGNSGGQGTTPRAVGRVVALQGTVOASDSLGTGAETLNGLIQFDAAIQPDSDSGPV 187
DB 153 VVMGNSGGQGTTPRAVGRVVALQGTVOASDSLGTGAETLNGLIQFDAAIQPDSDSGPV 212
QY 188 VVGLGVVGMNTAASDNFOLSGGGFAIPICQAVAIAGQIRSGGSGPTVHIGTAFGL 247
DB 213 VVGLGVVGMNTAASDNFOLSGGGFAIPICQAVAIAGQIRSGGSGPTVHIGTAFGL 272
QY 248 GVVDNNGGARVQVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHHPGDVIS 307
DB 273 GVVDNNGGARVQVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHHPGDVIS 332
QY 308 VTWQKSGGTRTGNVTLAEGPPA 330
DB 333 VTWQKSGGTRTGNVTLAEGPPA 355

RESULT 3
Q50320 PRELIMINARY; PRT; 361 AA.
ID Q50320 PRELIMINARY; PRT; 361 AA.
AC Q50320; (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE 34kDa protein precursor.
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JD89/107;
RX MEDLINE=95005449; PubMed=7921248;
RA Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
RT "Identification and characterisation of a putative serine protease
expressed in vivo by Mycobacterium avium subsp paratuberculosis.";
RL Microbiology 140:1977-1982(1994).
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR EMBL; Z23092; CAA80638.1; -.
DR PIR; S47170; S47170.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1c.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00106; PDZ; 1.

```

KW Hydrolase; Protease; Serine protease; Signal.
 FT SIGNAL 1 38 POTENTIAL.
 SQ SEQUENCE 361 AA; 35709 MW; 30FEF78FD6F3C411 CRC64;

Query Match 72.5%; Score 1231.5; DB 2; Length 361;
 Best Local Similarity 73.1%; Pred. No. 4.9e-59; Indels 1; Gaps 1;
 Matches 236; Conservative 38; Mismatches 48; Indels 1; Gaps 1;

QY 8 APPALSQDRFAFFPALPLDPSAMVAGVQPVNINIKLGYNNNAVAGTGIVIDPNGVLTN 67
 DB 40 APSGLALDFADRLAPIDPSAMVQGVQPVNIDTKFGYNNNAVAGTGIVIDPNGVLTN 99
 QY 68 NNHVIAGATDINAFSVGSQTVGVVDVVDRTQDVAVLQIRGAGLPSNAIGGGVAVGEP 127
 DB 100 NNEVISGATEISAFDVGNGQTVADVVDVVDRTQDVAVLQIRGAGLPSNAIGGGVAVGEP 159
 QY 128 VVAMGNSGGGTPPRAVGRVVALGTVQASDLSLTGAETTLNGLIOFADAAIOPGSGGPPV 187
 DB 160 IVALGVNGGGTPPRAVGRVVALGTVQASDLSLTGAETTLNGLIOFADAAIOPGSGGPPV 219
 QY 188 VNLGQVGVGNTAAASDNFOLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGLG 247
 DB 220 VNSAGQVIGVDTAATDSYKMS-GGQGFALPIGRAMAVANQIRSGAGSNVTHIGPTAFGLG 278
 QY 248 GVVDNNGGARVORVVGSAASLSTGSDVITAVDGPINSATAMADALNGHHFGDVIS 307
 DB 279 GVVDNNGGARVORVVGSAASLSTGSDVITAVDGPINSATAMADALNGHHFGDVIS 338
 QY 308 VTWQKSGGTRTGNVTLAEGPPA 330
 DB 339 VHFPSVDGGERTANITLAEGPPA 361

RESULT 4
 Q9CCY9 PRELIMINARY; PRT; 354 AA.
 ID AC Q9CCY9
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Probable secreted serine protease.
 GN ML2659.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RA Cole S.T., Biglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus."
 RL Nature 409:1007-1011(2001).
 CC -/- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL; AL583926; CAC32191.1; -;
 DR PIR; A87242; A87242.
 DR Lepnora; ML2659; -;
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR008256; Peptidase_S1B.
 DR InterPro; IPR001940; Peptidase_S1C.

Query Match 70.6%; Score 1198; DB 16; Length 354;
 Best Local Similarity 72.0%; Pred. No. 3.1e-57;
 Matches 232; Conservative 40; Mismatches 46; Indels 4; Gaps 2;

QY 9 PPALSQDRFAFFPALPLDPSAMVAGVQPVNINIKLGYNNNAVAGTGIVIDPNGVLTN 68
 DB 37 PSTLALDRFSRNPPLNPAANVA---PQVNNSTRLGYNNAVAGTGIVIDSSGVLTN 93
 QY 69 NNHVIAGATDINAFSVGSQTVGVVDVVDRTQDVAVLQIRGAGLPSAAIAGGGVAVGEPV 128
 DB 94 NNHVIAGATDINAFSVGSQTVGVVDVVDRTQDVAVLQIRGAGLPSAAIAGGGVAVGEPV 153
 QY 129 VAMGNSGGGTPPRAVGRVVALGTVQASDLSLTGAETTLNGLIOFADAAIOPGSGGPPV 188
 DB 154 VALGNTGGGGTPPRAVGRVVALGTVQASDLSLTGAETTLNGLIOFADAAIOPGSGGPPV 213
 QY 189 NGLGQVGVGNTAAASDNFOLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGLG 248
 DB 214 NSRGQVGVGNTAAADNXYM-LGGQGFALPIQAMAEVVGAIKSGAGSNVTHIGPTAFGLG 272
 QY 249 VVDNNGGARVORVVGSAASLSTGSDVITAVDGPINSATAMADALNGHHFGDVISV 308
 DB 273 VLDNNGGARVORVVGSAASLSTGSDVITAVDGPINSATAMADALNGHHFGDVISV 332
 QY 309 TWQKSGGTRTGNVTLAEGPPA 330
 DB 333 NYRSAGGGDLTANVTLAEGPPA 354

RESULT 5
 Q8VK44 PRELIMINARY; PRT; 446 AA.
 ID AC Q8VK44
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Heat shock protein Htra, putative.
 GN MT1011.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains";
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE006985; AAK45259.1; -;
 DR TIGR; MT1011; -;
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SMO0228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 DR SEQUENCE 446 AA; 44484 MW; 54170CBEA8F872B CRC64;

Query Match 27.1%; Score 460.5; DB 16; Length 446;
 Best Local Similarity 37.4%; Pred. No. 2.5e-17;
 Matches 122; Conservative 55; Mismatches 118; Indels 31; Gaps 10;

QY 21 PALPLDPSA---MVAQVGPOVNNIKLGYNAVAGTGVIDPVGVLNNHVIAGAT- 76
 DB 129 PAANPPGSEVQVAAKVVPVSVVMTDLGRQSE--EGSGIILSAEGLIITNNHVIATAAK 186

QY 77 -----DINAFVSGSQYGVYDRTQDVAVLQRLGAGGLPSAAIGGG--VAVGEP 127
 DB 187 PPLGSPPKTTVTFSDGRTAPFTVVGADPTSDIAVVRVQVSGLTPIISGSSDRLRVGP 246

QY 128 VVAMGNSGGQGTTPRAVGRVVALGQTVQASDSLTSAGETLNGLIQDAAIQPDGSGGPV 187
 DB 247 VLAIGSPGLEGT--VTTGIVSALNRPVSTTGE--AGNQNTVLDATQTDAAINPNSGGAL 303

QY 188 VNGLGQVVGMMNTA---ASDNFQLSQG--GQGFAPICQAMAIAGQIRSGGSGPTVHIGP 241
 DB 304 VVMAQLGVNSAIATLGDASADAQSGSIGLGFAPVQAKRIADELISTGKA-----S 357

QY 242 TAFGLGLGVV-DNNGNGARVQVGSAPASLSIGISGTVITAVDGAIPNSATAMADALNGH 300
 DB 358 HASLGVQVNTDKDTPGAKIVEVAGGAANAAGVPGVVTVKVDRPINSADALVAARSK 417

QY 301 HPGDVISVWTQKSGGRTGNVTLAE 326
 DB 418 AFGATVALTQDPGSGSRTVQVTLCK 443

RESULT 6
 QY002 ID Q7U0X2 PRELIMINARY; PRT; 464 AA.
 AC Q7U0X2;
 DT 01-OCT-2003 (TREMELrel. 25, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Probable serine protease (Serine proteinase) (EC 3.4.21.-).
 GN MB1009.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 CX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL; BX248337; CAD93870.1; -.
 KW Hydrolase; Complete proteome.
 SQ SEQUENCE 464 AA; 46436 MW; AE93A4BB3FFA9BE3 CRC64;

Query Match 27.4%; Score 460.5; DB 16; Length 464;
 Best Local Similarity 37.4%; Pred. No. 2.6e-17;
 Matches 122; Conservative 55; Mismatches 118; Indels 31; Gaps 10;

QY 21 PALPLDPSA---MVAQVGPOVNNIKLGYNAVAGTGVIDPVGVLNNHVIAGAT- 76
 DB 147 PAANPPGSEVQVAAKVVPVSVVMTDLGRQSE--EGSGIILSAEGLIITNNHVIATAAK 204

QY 77 -----DINAFVSGSQYGVYDRTQDVAVLQRLGAGGLPSAAIGGG--VAVGEP 127
 DB 205 PPLGSPPKTTVTFSDGRTAPFTVVGADPTSDIAVVRVQVSGLTPIISGSSDRLRVGP 264

QY 128 VVAMGNSGGQGTTPRAVGRVVALGQTVQASDSLTSAGETLNGLIQDAAIQPDGSGGPV 187
 DB 265 VLAIGSPGLEGT--VTTGIVSALNRPVSTTGE--AGNQNTVLDATQTDAAINPNSGGAL 321

QY 188 VNGLGQVVGMMNTA---ASDNFQLSQG--GQGFAPICQAMAIAGQIRSGGSGPTVHIGP 241
 DB 322 VVMAQLGVNSAIATLGDASADAQSGSIGLGFAPVQAKRIADELISTGKA-----S 375

QY 242 TAFGLGLGVV-DNNGNGARVQVGSAPASLSIGISGTVITAVDGAIPNSATAMADALNGH 300
 DB 376 HASLGVQVNTDKDTPGAKIVEVAGGAANAAGVPGVVTVKVDRPINSADALVAARSK 435

QY 301 HPGDVISVWTQKSGGRTGNVTLAE 326
 DB 436 AFGATVALTQDPGSGSRTVQVTLCK 461

RESULT 7
 OS3896 ID OS3896 PRELIMINARY; PRT; 464 AA.
 AC OS3896;
 DT 01-JUN-1998 (TREMELrel. 06, Created)
 DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Putative serine protease.
 GN RV0983 OR MT0044.11.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 CX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Brown T., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krohn A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence."
 RL Nature 393:537-544(1998).
 CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL; AL021999; CAA17582.1; -.
 DR PIR; C70821; C70821.
 DR Tuberculin; RV0983; -.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SMO0228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KW Hydrolase; Protease; Serine protease; Complete proteome.
 SQ SEQUENCE 464 AA; 46452 MW; AB93BFC53E1EC8F CRC64;

Query Match 27.1%; Score 459.5; DB 16; Length 464;
 Best Local Similarity 37.4%; Pred. No. 2.9e-17;
 Matches 122; Conservative 55; Mismatches 118; Indels 31; Gaps 10;

QY 21 PALPLDPSA---MVAQVGPOVNNIKLGYNAVAGTGVIDPVGVLNNHVIAGAT- 76
 DB 147 PAANPPGSEVQVAAKVVPVSVVMTDLGRQSE--EGSGIILSAEGLIITNNHVIATAAK 204

QY 77 -----DINARVSGGQTYGVVDVYDRTQDVAVLQLEAGGLPSAAIGGG--VAVGEP 127
Db 205 PPLGSPPPKTTVTFTSDRTAPTFTWGADEFTSDIAVVRVQVSGLTPISLGSSDLRVQGP 264
QY 128 VVAMNGSGGGGPPRAVPGRVVAGLGTQVQASDLSLTGAETLNGLIQFDAAIOPGSGGPPV 187
Db 265 VLAIGSPGLEGT--VTTGIVSALNRPVSTTGE--AGNQNTVLDALQTDAAINPGNSGGAL 321
QY 188 VNLGLGVVGNMNTA-----ASDNFQLSG--GGGFAIPIGQAMAIAGIRSGGSPVTHIGP 241
Db 322 VNMNAQLVGVNSAIATLGDADSDAQAQSSIGLGFAPVDAQRIADELISTGKA-----S 375
QY 242 TAFGLGVV--DNNNGARVQRVVGVSAPASLGLSTGDVITAVDGAIPNSATAMADALNGH 300
Db 376 HASLGVQVNDKDTLGAKEVVEVAGGAANAGVPGKVVVTKVDDRPINSADALVAARVSK 435
QY 301 HFGDVISVTWQTKSGGTRTGNVTLAE 326
Db 436 APGATVALTQDPDSGGSRVTQVTLGK 461

RESULT 8
Q9CD67
ID Q9CD67 PRELIMINARY; PRT; 382 AA.
AC Q9CD67;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Possible secreted serine protease.
GN ML0176.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mangall K., Braham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Dutcoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL583917; CAC29684.1; -
DR PIR; H86930; H86930.
DR Leproma; ML0176; -
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 382 AA; 37084 MW; 3DD8DD8AE32A80D CRC64;

Query Match 26.5%; Score 450; DB 16; Length 382;
Best Local Similarity 37.5%; Pred. No. 7.5e-17;
Matches 126; Conservative 52; Mismatches 116; Indels 42; Gaps 12;
QY 18 ADFPALPLDPSAMVAQV----PQVNVNIN*KLGVNNAVGAGTGIVIDPNQWLTNNHVA 73

Db 59 ASVPAANM--FGSVEQVAVKVVSVVWLETDLGRQSE--EGSGVILSADGLILTNHVA 115
QY 74 GA-----TDINAFVSGGQTYGVVDVYDRTQDVAVLQLEAGGLPSAAIG 119
Db 116 VAAKPGGGPGGLSPKTTVTFTF---DGRASFTVVGADFTSDIAVVRVQVSGLTPITMG 172
QY 120 GG--VAVGEPVWAMNGSGGGGTPRAVPGRVVAGLGTQVQASDLSLTGAETLNGLIQFDAA 177
Db 173 SSADLRVGPQVAVVAGSPGLAGT--VTSGIVSALNRPVSTTGE--SGNQNTVLDALQTDAA 229
QY 178 IQPGSDSGPPVNVGLGVVGMNTAA-----SDNFQLSGGGGFAIPIGQAMAIAGIRSG 231
Db 230 INPNSGGALVNMGGQLVGVNSAIALTGADSDAQAQSSIGLGFAPVDAQRIADELIST 289
QY 232 CGSTVTHIGTFAFLGLGVVDNNGN--GARVQRVVGVSAPASLGLSTGDVITAVDGAIPNSA 290
Db 290 G--KATH-----ASLGVQVATDKGTPGAKVMDVWAGGAANAAPKGVVLTAKVDDRLISSA 343
QY 291 TAMADALNGHHPGDVIVSVTWQTKSGGTRTGNVTLAE 326
Db 344 DALVAARVSKAPGDKVSLTYQDSGSRVTQVTLGK 379

RESULT 9
Q9Z5G6
ID Q9Z5G6 PRELIMINARY; PRT; 452 AA.
AC Q9Z5G6;
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Putative serine protease.
GN MLCB373.28.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RL Harris D., Taylor K.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=93188700; PubMed=8446027;
RA Eigmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT "Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae."
RL Mol. Microbiol. 7:197-206(1993).
DR EMBL; AL035500; CAB36690.1; -
DR PIR; T45448; T45448.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 452 AA; 45130 MW; 6CA675EB0911F983 CRC64;

Query Match 26.5%; Score 450; DB 2; Length 452;
Best Local Similarity 37.5%; Pred. No. 9.3e-17;
Matches 126; Conservative 52; Mismatches 116; Indels 42; Gaps 12;

QY 18 ADFFPALPLDPSAMVAQVQ-----POVNNINTKLGYNNVAGAGTGIVDPNGWLTNNHVA 73
 Db 129 ASVPAANM-PSGSVEQAVKVPVSWLELTLGRQSE--EGSGVILSADGLILTNHVA 185
 QY 74 GA-----TDINAFSGSQTVGVVGVVDRDQVAVLQRLGAGGLPSAIG 119
 Db 186 VAAKPGGGGGLSPKTVTFP-----DGTASTFTVVGADPTSDIAVVRVQSGLTITMG 242
 QY 120 GG--VAVGEPVAMNGSGGGGTTPRAVPGRVVAGLQTVQASDSLTAETLNGLIQFDDA 177
 Db 243 SSADLRVGQPVVAVGSPGLAGT--VTSGIVSALNRVPSTGTB-SGNQNTVLDAITDAA 299
 QY 178 IQPGDSGGPVVNGLGQVVGWNTAA-----SDNFQLSQGGGPAIPGQMAIAGIRSG 231
 Db 300 INPNSGGALVNMGGQLVGVNSAIALTGADSGDAQSGSIGLGAIPVDAQRIADELIST 359
 QY 232 GGSPTVHIGPTAFGLGLGVVWNGN-GARVQVVGSGAPASLIGISTGDIVTAVDGPINSA 290
 Db 360 G--KATH-----ASLGQVVAIDKGPFGAKVMDVWAGGAANAAPVKGVLLTKVDDRLISSA 413
 QY 291 TAMADALNHHHPGPDVIVSVTWQTSKSGTRTGNVTIAE 326
 Db 414 DALVAAVRSKAPGDKVSLTYQDQSGSKRTVQVTLGK 449

RESULT 10
 Q93J30 ID Q93J30 PRELIMINARY; PRT; 519 AA.
 AC Q93J30;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative protease.
 GN SC03977 OR SCBAC25E3.14.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Collins M., Harris D.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Redenbach M., Kieser H.M., Denapáite D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RL "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000983;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 Harper D., Batenan A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Horneby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RL "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RT Nature 417:141-147(2002).
 RL Nature 417:141-147(2002).

DR EMBL; AL939118; CAC44701.1; -.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR Hydrolase; Protease; Serine protease; Complete proteome.
 KW Hydrolyase; Protease; Serine protease; Complete proteome.
 SQ SEQUENCE 519 AA; 50327 MW; 5CB8D5F0CC19E428 CRC64;

Query Match 23.4%; Score 397.5; DB 16; Length 519;
 Best Local Similarity 34.4%; Pred. No. 7.5e-14;
 Matches 109; Conservative 51; Mismatches 128; Indels 29; Gaps 11;

QY 28 SAMVAQVGPQVNNINTKLGYNNVAGAGTGIVDPNGWLTNNHVAIAGATDINAFSV--GS 85
 Db 205 AGVAAKALPSTVTIQAE--GSNGEGGTGTGFVFDKEGHIVTNNHVAEAVDGGKLSATFPN 263
 QY 86 QQTVGVVGVVDRDQVAVLQRLGA--GGLPSAAIAGG--VAVGEPVAMNGSGGGGTTPR 142
 Db 264 GKXYDAEYVGHQAQYGVAVIKLENAPSLDKPLALGDSKVAVGDSTTAIGAPFLSNT-- 321
 QY 143 AVPRGVALGQTVQASDSLTAETLNGLIQFDDAIIOPGDSGPGVAVNGLGQVVGWNTA-- 200
 Db 322 VTTGIISAKRPVASSDGSADSKASYMSALQTDASINPNSGGPFLDQAQNVIGINSAIQ 381
 QY 201 --ASDNFQLSQGGQ---GFAIPICQAMAIAGIRSGSGSPVTHIGPTAFGLGVVWNGN 255
 Db 382 STNGGGFGTQAGSIGLGAIPVNAQKFAVQQLIKSGKPVVAKIGASVSL-----EETT 436
 QY 256 GARV--QRVVGSG-----APASLIGISTGDIVTAVDGPINSAIATADALNHHHPGDIS 307
 Db 437 GAKUTEQGVGSDPVEKGPADDAGLKPGDVITKLDREVDSGPTLIGETWTHKPGDEV 496
 QY 308 VTWQTSKSGTRTGNVT 324
 Db 497 VTVE--RGKQHTAEVTL 512

RESULT 11
 Q8DG87 ID Q8DG87 PRELIMINARY; PRT; 375 AA.
 AC Q8DG87;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Serine protease.
 GN TLR2436.
 OS Synecococcus elongatus (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
 OX NCBI_TaxID=32046;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1;
 RX MEDLINE=22251544; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Watanabe A., Iiziguchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
 RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RL "Complete genome structure of the thermophilic cyanobacterium
 Thermosynechococcus elongatus BP-1.";
 DNA Res. 9:123-130(2002).
 DR EMBL; AF005377; BAC0988.1; -.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=33903;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RA "Genome sequence of an industrial microorganism Streptomyces
 RT avermitilis: deducing the ability of producing secondary
 RT metabolites.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hattori M., Omura S.;
 RA "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomyces avermitilis.";
 RL Nat. Biotechnol. 21:526-531(2003).
 DR EMBL; AP005038; BAC71935.1; -
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR003003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR Protease; Complete proteome.
 KW SEQUENCE 472 AA; 4086 MW; AE04FAA09A3B7E0 CRC64;

Query Match 22.6%; Score 384; DB 16; Length 472;
 Best Local Similarity 32.5%; Pred. No. 3.6e-13;
 Matches 103; Conservative 57; Mismatches 133; Indels 24; Gaps 10;

QY 28 SAMVAQVGPQVNTKLGYNNAVAGTGIVDPNGVLTNNHVIAGATDINAFSV--GS 85
 Db 157 AGVAAKALPTVTIEAQSSSEG-GTGTGTFVFDKQGHIVTNHVAEAVDGGKLTATFPD 215
 QY 86 GQTYGVDDVGYDRTQDVAVLQLRGA--GGLPSAAIGGG--VAVGEPVVMNGSGGGGTPT 142
 Db 216 GKYNAEVVGHAQGYDVAVVVKLNAPSDLQLPLGDSKVAVGDSITAIAGAPFLSNT-- 273
 QY 143 AVGRVVALQQTVAQSDSLTGAETLNGLIQFDAAIQPGDSGPPVNGVGVGWNMTA-- 200
 Db 274 VTITGIISAKRNPVASSDSSSSKASTYMSALQTDASINPGSGPLDQAQSGVIGNSAIQ 333
 QY 201 ASDNQLSQGGQ-----GFAIPGQAMAIAGQIRSGGSPVTHIGPTAFL-----GLGV 249
 Db 334 SSSSGGLSGSGSGSGLGLFAIPINAKYVAQELIKTPVPVIGASVLEEGTGAKI 393
 QY 250 VDNNNGARVQRVVGAPASLGISTGVDITAVDGPAPINSATAMADALNGHPGVDISVT 309
 Db 394 TEOGASGS--DAITPNPAAKAGKPGDVITKLDMDVIDSGPTLIGEIMWTRPGATVKLT 451
 QY 310 WQTKSGTGRIGNVTLAE 326
 Db 452 Y-TRDKARTDVTLLGE 467

RESULT 14

Q821L8 PRELIMINARY; PRT; 619 AA.
 AC Q821L8;

DT 01-JUN-2003 (TEMBLrel. 24, Created)

DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Putative serine proteinase.
 GN SAV3115.
 OS Streptomyces avermitilis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RA "Genome sequence of an industrial microorganism Streptomyces
 RT avermitilis: deducing the ability of producing secondary
 RT metabolites.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hattori M., Omura S.;
 RA "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomyces avermitilis.";
 RL Nat. Biotechnol. 21:526-531(2003).
 DR EMBL; AP005033; BAC70826.1; -
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SMO0228; PDZ; 1.
 KW Complete proteome.
 SQ SEQUENCE 619 AA; 64251 MW; D2D3C72A67D3B3A6 CRC64;

Query Match 22.6%; Score 384; DB 16; Length 619;
 Best Local Similarity 34.0%; Pred. No. 4.9e-13;
 Matches 115; Conservative 53; Mismatches 126; Indels 44; Gaps 12;

QY 13 SQDRFADFPALPLDPSAMVAQVGPQVNTKLGYNNAVAGTGIVDPNGVLTNNHVI 72
 Db 298 SEERAADSV-----GIAARALPSVVTLVHK--GSAAGTGTGFLDGRGHILTNNHV 349
 QY 73 --AGATDINAFSGSGTGYGVVDYDRTQDVAVLQLRGAGGLPSAAIGG--GVAVGEPV 128
 Db 350 EPAGSSGEISVTFSGGTAKATVVGGRDSGYDLAVVKVSGVGLKMPGLGNSDNVQGDVP 409
 QY 129 VAMNGSGQGQGTTPRAGRVRVVALGQTQCA--SDSLTGAETLNGLIQFDAAIQPGDSGPPV 187
 Db 410 VAIGAPDLANT--VTSGIISAKERPITAGEKGDGSDVSVVALQTDAPINFGNSGGLP 467
 QY 188 VNLGQVGVGWNMTA-----SDNQLSQGGQ---GFAIPGQAMAIAGQIRSGGSPVTHIG 240
 Db 468 LDSKARVGVINSAIRSADSSDQSGAGSIGLGLFAIPVNAQKVAEELINTGRATHPVIG 527
 QY 241 PTAFLGLGVDDNNGGARV-----QRVVGSAASLGISTGVDITAVDGPAPINSATA 292
 Db 528 VTLDM-----DYTGDGARVGTKNDGSGSPVTRGGPDRAIGQAGDVITTEVDGQRHSGE 582
 QY 293 MADALNGHFGDVISVTWQTKSGGTRTGN---VTLAEG 327
 Db 583 LIVKIRAHKPDRLALVE-----RDGKEKPVTLVLG 614

RESULT 15

Q8FRI7

ID	Q8FR17	PRELIMINARY;	PRT;	473 AA.
AC	Q8FR17			
DT	01-MAR-2003	(TrEMBLrel. 23, Created)		
DT	01-MAR-2003	(TrEMBLrel. 23, Last sequence update)		
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)		
DE	Putative serine protease, heat shock protein.			
GN	CE0950.			
OS	Corynebacterium efficiens.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Corynebacteriaceae; Corynebacterium.			
OX	NCBI_TaxID=152794;			
RN	[1] SEQUENCE FROM N.A.			
RP	STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;			
RC	Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,			
RA	Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,			
RA	Usuda Y., Sugimoto S.;			
RT	The entire genomic sequence of Corynebacterium efficiens YS-314.;			
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF003217; BAC17760.1; -.			
DR	GO; GO:0003773; F:heat shock protein activity; IEA.			
DR	GO; GO:0008233; F:peptidase activity; IEA.			
DR	GO; GO:0004295; F:trypsin activity; IEA.			
DR	GO; GO:007242; P:intracellular signaling cascade; IEA.			
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR	InterPro; IPR009003; Cys_Ser_trypsin.			
DR	InterPro; IPR001478; PDZ.			
DR	InterPro; IPR001254; Peptidase_S1.			
DR	InterPro; IPR001940; Peptidase_S1C.			
DR	Pfam; PF00595; PDZ; 1.			
DR	Pfam; PF00089; trypsin; 1.			
DR	PRINTS; PR00834; PROTEASES2C.			
DR	SMART; SM00228; PDZ; 1.			
DR	PROSITE; PS50106; PDZ; 1.			
KW	Heat shock; Protease; Complete proteome.			
QY	SEQUENCE 473 AA; 47757 MW; 5C0C6C9561F8C524 CRC64;			
QY	Query Match	22.3%;	Score 379;	DB 16; Length 473;
Db	Best Local Similarity	34.3%;	Pred. No. 6.6e-13;	
Db	Matches 108; Conservative	50;	Mismatches 131; Indels 26; Gaps 9;	
Qy	22	ALPLDPSAMVAQYQPVNINTKLGNNAVAGTGIVDPNGVVLNNHVIAGATDINAF	81	
Db	170	AEFGSAEQVADQVLPSVVSQALT--RNSASEGSGSISSDGFVMTNNHVVAGVEQNGVL	227	
Qy	82	SV--GSGQTYGVVGVYDRTQDVAVLQLEGAGGLPSAAIG--GGVAVGEPVVMGNSGGQ	137	
Db	228	QVTWSDGSSYEADFIAGDVSTDIATKIRNATLPMINFGDSDSLAVGQEVLA VGSPLGL	287	
Qy	138	GGTPRAVPGRVVALGQTVQASDSLTGAETINGLIQFDAAIQCDSDGPGVYVNGLGQVVM	197	
Db	288	SST--VTSIGIVSALNRPVRASGD--GGESSLIDAIQTDAAINFGNSGGPLVDMEGNLIGM	343	
Qy	198	NTAASDNFQLSQ-----GQGFAPIGQAMAIAGQIRSGGSPVTHIGTAFGLGVV	250	
Db	344	NSVIA---SISGTDTCGSGIGLGFAPANFAKRVATQLIETGVATQPMIGVTLANGANV-	399	
Qy	251	DNNGNGARVQVVGSAAPASIGSTGVITAVDGPATNSATAMADALNHHHPGDVISVTW	310	
Db	400	----SGALIASVQGGFGAAGLXPGDVVTMLNDRVIDSDALIAAVRSQDFGETVTLTV	455	
Qy	311	QTK-SGGTRTGNVTL 324		
Db	456	TAQDTSQTRQVEVTL 470		

Search completed: June 22, 2004, 17:22:46
Job time : 27.9826 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:46:13 ; Search time 5.49204 Seconds
(without alignments)
3128.737 Million cell updates/sec

Title: US-09-886-349A-6
Perfect score: 1698
Sequence: 1 MHHHHHAPPALSQDRPADF.....QTKSGTGTGNVLAEGPPA 330

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	323.5	19.1	458	1 YVTA_BACSU	Q9r9i1 bacillus su
2	323	19.0	413	1 HTRA_LACHE	Q9z4h7 lactobacil
3	316	18.6	355	1 DEGS_ECOLI	P31137 escherichia
4	312	18.4	408	1 HTRA_LACLA	Q91a06 lactococcus
5	305.5	18.0	513	1 DEGP_BRUME	Q8Y932 bruceella me
6	304.5	17.9	513	1 DEGP_BRUSU	Q44597 bruceella su
7	302.5	17.8	455	1 DEGO_ECOLI	P39099 escherichia
8	293	17.3	475	1 DEGP_SALTY	P26982 salmonella
9	293	17.3	504	1 DEGP_RHIME	Q52894 rhizobium m
10	291.5	17.2	449	1 HTRA_BACSU	Q34358 bacillus su
11	291.5	17.2	474	1 DEGP_ECOLI	P09376 escherichia
12	289	17.0	466	1 HTOA_HAEIN	P45129 haemophilus
13	285.5	16.8	503	1 DEGP_BAHE	P54925 bartonella
14	285	16.8	497	1 DEGP_CHLTH	P18584 chlamydia t
15	284	16.7	478	1 DEGP_BUCAL	P57322 buchnera ap
16	280.5	16.5	497	1 DEGP_CHLMO	C9p197 chlamydia m
17	279	16.4	488	1 DEGP_CHLPN	Q9z260 chlamydia p
18	278	16.4	437	1 DEGI_ARATH	C22609 arabidopsis
19	277	16.3	508	1 DEGS_RICCN	Q927a1 rickettsia
20	271.5	16.0	400	1 VYXA_BACSU	P19568 bacillus su
21	269.5	15.9	448	1 DEGS_ARATH	Q91u10 arabidopsis
22	267.5	15.8	478	1 DEGP_BUCAP	C08591 buchnera ap
23	251	14.8	513	1 DEGP_RICPR	O05942 rickettsia
24	245.5	14.5	458	1 HRA2_HUMAN	Q43464 homo sapien
25	239	14.1	340	1 DEGS_HAEIN	P44947 haemophilus
26	232.5	13.7	480	1 HRA1_HUMAN	Q92743 homo sapien
27	231.5	13.6	453	1 HRA3_HUMAN	P83110 homo sapien
28	225.5	13.3	458	1 HRA2_MOUSE	Q9j1y5 mus musculu
29	222	13.1	476	1 HRA4_HUMAN	P83105 homo sapien
30	218.5	12.9	460	1 HRA3_MOUSE	Q9d336 mus musculu
31	218	12.8	480	1 HRA1_MOUSE	Q9r118 mus musculu
32	175.5	10.3	321	1 SPFA_ARATH	Q9se17 arabidopsis
33	172.5	10.2	630	1 Y4BJ_RHISN	P55377 rhizobium s

ALIGNMENTS

RESULT 1

ID	YVTA_BACSU	STANDARD;	PRT;	458 AA.
AC	Q9r9i1: Q35021; Q35039;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Probable serine protease yvta (EC 3.4.21.-).			
GN	YVTA OR YVTB OR BSU33000.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=20158875; PubMed=10692364;			
RA	Noone D., Howell A., Devine K.M.;			
RT	"Expression of ykda, encoding a Bacillus subtilis homologue of Htra,			
RT	is heat shock inducible and negatively autoregulated";			
RL	J. Bacteriol. 182:1592-1599(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98015415; PubMed=9353931;			
RA	Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;			
RT	"Sequencing of regions downstream of addA (98 degrees) and citG (289			
RT	degrees) in Bacillus subtilis";			
RL	Microbiology 143:3305-3308(1997).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Borriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,			
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,			
RA	Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,			
RA	Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,			
RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,			
RA	Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,			
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,			
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,			
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,			
RA	Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,			
RA	Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,			
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,			
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,			
RA	Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,			
RA	Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,			
RA	Takeuchi M., Tamakoehi A., Tanaka T., Terpsstra P., Tognoni A.,			
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,			

Q64761 avian adeno
P04985 bos taurus
Q50594 mycobacteri
Q99372 rattus norv
P52320 scryptomyce
P56877 mycobacteri
Q53810 mycobacteri
Q06794 mycobacteri
Q52657 rickettsia
P12021 sus scrofa
C53047 r outer mem
Q06653 r outer mem

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; AE009571; AAL52511.1; --

PIR; AD3418; AD3418.

InterPro; IPR003005; Cys_Ser_trypsin.

InterPro; IPR001478; PDZ.

InterPro; IPR001254; Peptidase_S1.

InterPro; IPR001940; Peptidase_S1C.

Pfam; PF00595; PDZ; 2.

Pfam; PF00089; trypsin; 1.

PRINTS; PR00834; PROTEASES2C.

SMART; SM02228; PDZ; 2.

PROSITE; PS0106; PDZ; 2.

Hydrolase; Serine protease; Periplasmic; Repeat; Signal; Complete proteome.

KW SIGNAL 1 25

FT CHAIN 26 513

FT DOMAIN 125 299

FT DOMAIN 300 391

FT DOMAIN 414 500

FT ACT_SITE 152 152

FT ACT_SITE 182 182

FT ACT_SITE 257 257

SQ SEQUENCE 513 AA; 53514 MW; C472FEF9DFC6268 CRC64;

Query Match 18.0%; Score 305.5; DB 1; Length 513;

Best Local Similarity 34.0%; Pred. No. 1e-10;

Matches 103; Conservative 37; Mismatches 118; Indels 45; Gaps 13;

QY 46 GYNNAVGAGTGVIDPNEGVLTNHNVAGATDINAFSV--GSGQTYGVVDVGYDRTQDVA 103

DB 128 GHERPVAQSGGFVISEDGVYVNNHVV--SDGDAYVLLDDGTDLAKLIGADPRTDLA 184

QY 104 VLOIRAGGLPS-----AATG--GGVAVGEPVWAMNGSGGGTTPRAVGRVVALGQTQV 156

DB 185 VLKINA----PKRFVTVAFGDDNKNRVGDMVAVGNPFLGGT--VTSIGIVSARGRDIG 238

QY 157 ASDSLTCABETLNLGIQDAAIQPDAGGPPVWGLGVGMVGMNTAASDNFQLSGQGQ--F 214

DB 239 AG-----PYDDFLQIDAAVKNKSGGFADLSGEVIGINTAI---FSPSGSGVGIAP 287

QY 215 AIPQOAWAAGQIRSGG----GSPTVHIGP-----TAFGLGVVDNNGARVORVVGVS 265

DB 288 AIPSTAKVQVDLIKKGSVERGWIGVQIQPVTKDIAASGLA---EEKGAIVASPODD 343

QY 266 APAASLGISGDVITAVDGPAPINCATAMADALNGHPCDDVISVTWQKSGGTRIGNVTLA 325

DB 344 GPAAKGIKAGDVITAVNGETVQDPRLARKVANIAPGEKAALIVWRKNKAEI-NVTIIA 402

QY 326 EGP 328

DB 403 AMP 405

RESULT 6

DEGP BRUSU STANDARD; PRT; 513 AA.

ID DEGP BRUSU

AC Q44597;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Probable serine protease do-like precursor (EC 3.4.21.-).

GN DEGP OR HTRA OR BR0611.

CS Brucella suis, and

OS Brucella abortus.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Brucellaceae; Brucella.

OX NCBI_taxid=29461, 235;

RN [1]

RC SEQUENCE FROM N.A.

RC SPECIES=B.abortus; STRAIN=2308;

Db 288 AIFSTAKQVVQDLIKKSGVERGWIGVQIQVTKDIAASLGLA-----BEKGAIVASPOD 343
 QY 266 APAASLGISTGVDITAVDGAIPINSATAMADALNGHHPGDVIVTWTQKSGGRTGNVTIA 325
 Db 344 GPAKAGIKAGDVITAVNGTVDQPRDLARKVANIAPGEKAALTWRKKASEI-NVTIA 402
 QY 326 EGP 328
 Db 403 AMP 405

RESULT 7
 DEGP_ECOLI STANDARD; PRT; 455 AA.
 AC P39099;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Protease degQ precursor (EC 3.4.21.-).
 GN DEGP OR HHOA OR B3234.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RA Bass S., Gu Q., Goddard A.;
 RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-32.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=96165272; PubMed=8576051;
 RA Waller P.R., Sauer R.T.;
 RT "Characterization of degQ and degS, Escherichia coli genes encoding
 RT homologs of the DegP protease."
 J. Bacteriol. 178:1145-1153(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=9742617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 Science 277:1453-1474(1997).
 RL
 CC -1- FUNCTION: PROTEASE WITH A SHARED SPECIFICITY WITH DEGP/DEGP.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- SIMILARITY: Belongs to peptidase family S2C.
 CC -1- SIMILARITY: Contains 2 PDZ/DHR domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U15661; AAC3992.1; -;
 CC EMBL; U32495; AAC44005.1; -;
 CC EMBL; U18997; AAA58036.1; -;
 CC EMBL; AE00402; AAC76266.1; -;
 CC PUR; JC6051; JC6051.
 CC MERO2S; S01.274; -;
 CC SWISS-2DPAGE; P39099; COLI.
 CC EcoGene; EG12612; degQ.
 CC InterPro; IPR009003; Cys_Ser_trypsin.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR001254; Peptidase_S1.
 CC InterPro; IPR001940; Peptidase_S1C.
 CC Pfam; PF00595; PDZ; 2.

DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS0106; PDZ; 2.
 KW Hydrolase; Serine protease; Periplasmic; Repeat; Signal;
 KW Complete proteome.
 FT SIGNAL 1 27
 FT CHAIN 28 455
 FT DOMAIN 258 349
 FT DOMAIN 355 447
 FT ACT_SITE 109 109
 FT ACT_SITE 139 139
 FT ACT_SITE 214 214
 FT ACT_SITE 214 214
 SQ SEQUENCE 455 AA; 47205 MW; 6A090F93AC021C83 CRC64;
 Query Match 17.8%; Score 302.5; DB 1; Length 455;
 Best Local Similarity 29.4%; Pred. No. 1.4e-10;
 Matches 100; Conservative 50; Mismatches 101; Indels 89; Gaps 14;
 QY 15 DRFADFFALPLDPSAMVAQVGPQVNVNITKLGYNNAVAGTGVIDEN-GVLTNNHVA 73
 Db 78 DDLDPQAPPE-----GLGSGVIINASKGVLTNNHVN 112
 QY 74 GATDINAFVSGSGQTVGVVGVYDRTQDVAVLQURGAGGLPSMAIGGG--VAVGEPVVM 131
 Db 113 QAQKI-SIQNDGREFDAKLIGSDQSDIALQLQNFPSKLTQIAIADSKLRVGDFAVAV 171
 QY 132 GNSGGQGGTPRAVGRVVALGVQTVQASDSLTGAE-ETLNGLIQPDAAIQPDAGGPVNG 190
 Db 172 GNPFGLGQT--ATSGIVSALGR-----SGLNLEGLENFQTDASINRGSGGALLNL 221
 QY 191 LGQVVGVNNTAASDNFOLSOGQ---GFAIPIGOAMAIA-----GOIRSGSGSPVTHIG 240
 Db 222 NGELIGNTAI-----LAPGGSGVIGFALPSPNWTALQQLIDFGEIKKG----- 267
 QY 241 PTAFLGLGVVDNNGN-----GARVQRVVGSSAPAAASLGISTGVDITAVDGAIPINS 289
 Db 268 ---LLGIGKTEMSADIAKAFNLDVQRCGAFVSEVLPGSGSAKAGVKAGDIITSLNGKFLNS 324
 QY 290 ATAMADALNGHHPGDVIVTWTQKSGGTRTG-----NVTIL 324
 Db 325 FAELRSRIATTEPGT-----KVKLGLLRNGKPLEVEVTL 358

RESULT 8
 DEGP_SALTY STANDARD; PRT; 475 AA.
 ID DEGP_SALTY
 AC P26982;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Protease do precursor (EC 3.4.21.-).
 GN DEGP OR HTRA OR PTD OR STM0209.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C5;
 RX MEDLINE=91251770; PubMed=1645840;
 RA Johnson K., Charles I., Dougan G., Pickard D., O'Gaora P., Costa G.,
 RA Ali T., Miller I., Hortaech C.;
 RT "The role of a stress-response protein in Salmonella typhimurium
 RT virulence."
 RL Mol. Microbiol. 5:401-407(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,


```
FT FT CONFLICT 464 504 KTWTSPTTSSPART (IN REF. 1).
FT FT KSDADDVLKVINNAKDRSKALFOIBAQSGSRFVALPITQG
FT FT -> NRQTTFSR (IN REF. 1).
SQ SEQUENCE 504 AA; 53035 MW; D7E82BB981EA23C CRC64;

Query Match 17.3%; Score 293; DB 1; Length 504;
Best Local Similarity 33.3%; Pred. No. 5.3e-10;
Matches 97; Conservative 40; Mismatches 116; Indels 38; Gaps 12;

Qy 54 GTGIVDPNGVLTNNHVIAGADINAFSV--GSGQTYGVVDVGYDRTQDVAVLQL--RG 109
Db 124 GSGFFITDGYLTNNHV---SDGSAFTVMKDGTELDKLVGKDSRTDLAVLKVDKDR 180
Qy 110 AGGLPSAAGGVAVGPVVMNGSGGGTTPRAVPGRVVVALGQTVQASDLSLTGAETLN 169
Db 181 KFTVVSFADDEKVRVGVDMVAVGNPFLGCTVTA--GIISARGRDIGSG-----PYD 230
Qy 170 GLIQFDAAIQPDAGGVPVNGLVGVGMNTAASDNFQLSGGGG--FAPIGQAMAIAGQ 227
Db 231 DYQVDAVAVRNGSGGTFNLSGEVGVINTAI---FSPGCGNVGIAFAIPASVAKDVVD 287
Qy 228 IRSGG-----GSPVTHIGPTA-----FLGLGVDPNNGNGARVQVRVWGSAPAASLGISTGV 278
Db 288 LKDGTVSRGWLGVQIQPVTKDIAESLGL-----SEANGALVVEPQAGSPQERAGIKNGDV 343
Qy 279 ITAVDGNPINSATAMADALNGHHPGDVIVT--WTKSGGTRTGNTVTLABCP 328
Db 344 VTALNGEPVXPDRLARVAALRPGSTAETVLTW--RSKSEVTNLEIGTLP 392

RESULT 10
HTRA_BACSU STANDARD; PRT; 449 AA.
ID_HTRA_BACSU STANDARD; PRT; 449 AA.
AC O34358;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable serine protease do-like htraA (EC 3.4.21.-).
GN HTRA OR BSU12903.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=168;
RA Devine K.M.;
RT "Sequence of the Bacillus subtilis genome between xlyA and ykoR.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabre C., Ferrari E., Foulger D.,
RA Fritze C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haeche J., Harwood C.R., Henaut A.,
RA Halbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega E., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Pressecan E., Pujic B., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield P.,
```

```
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vanderbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler B., Wedler H., Weitzneger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
[3]
RN TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=20158875; PubMed=10692364;
RA Noone D., Howell A., Devine K.M.;
RT "Expression of ykda, encoding a Bacillus subtilis homologue of HtrA,
RT is heat shock inducible and negatively autoregulated.";
RL J. Bacteriol. 182:1592-1599(2000).
[4]
RN TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=20576168; PubMed=1133960;
RA Noone D., Howell A., Collier R., Devine K.M.;
RT "Ykda and yvta, HtrA-like serine proteases in Bacillus subtilis,
RT engage in negative autoregulation and reciprocal cross-regulation of
RT ykda and yvta gene expression.";
RL J. Bacteriol. 183:654-663(2001).
[5]
RN TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=21439741; PubMed=1155295;
RA Hyvrylaenen H.-L., Bolhuis A., Darmon E., Muukkonen L., Koski P.,
RA Vitikainen M., Sarvas M., Pragai Z., Bron S., van Dijk J.M.,
RA Kontinen V.P.;
RT "A novel two-component regulatory system in Bacillus subtilis for the
RT survival of severe secretion stress.";
RL Mol. Microbiol. 41:1159-1172(2001).
CC -!- FUNCTION: May be involved in processing, maturation, or secretion
CC of extracellular enzymes.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- INDUCTION: Transcription is cssS dependent. Induced by heat shock
CC during exponential growth and by heterologous amylases at the
CC transition phase of the growth cycle. Negatively regulates its own
CC expression during exponential growth and during heat shock.
CC -!- MISCELLANEOUS: Inactivation results in compensating overexpression
CC of yvta, especially during stress conditions.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AJ002571; CAA05570.1; -
DR EMBL; Z99110; CAB13147.1; -
DR MEROPS; S01_273; -
DR Subtilist; BG12608; htra_Ser_trypsin.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1c.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR Hydrolase; Protease; Serine protease; Heat shock; Transmembrane;
KW Complete proteome.
```

FT DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 45 67 POTENTIAL.
FT DOMAIN 68 449 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 103 108 POLY-SER.
FT DOMAIN 146 152 POLY-SER.
FT DOMAIN 348 437 POZ.
FT ACT_SITE 179 179 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 209 209 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 290 290 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 449 AA; 47713 MW; E12B07A9018EE414 CRC64;
Query Match 17.2%; Score 291.5; DB 1; Length 449;
Best Local Similarity 29.0%; Pred.No.5.7e-10;
Matches 92; Conservative 53; Mismatches 127; Indels 45; Gaps 11;
QY 28 SAMVAQVGFQVNNINIKLGNNA-----VAGTGTGVI-DPENG--VVLINN 69
Db 119 SDWVEDLSPAIVGITNLQAGSSSLFGSSSDSDSTSGSGGVIFKXKNGKAVIINN 178
QY 70 HVIAGATDINARVSGSGQYGVVDVYQDRTQVAVLQLRAGAGLPSAIG--GGVAVGEP 127
Db 179 HVVEGASSLKV-SLYDGTETVAKLVGSDSLTDLAVLQISDDHVTKVANFGSDSLRTGET 237
QY 128 VVAMGNSGGQGTTPRAVPGRVVAGVQVSGVSLTGAERTNLGLQFDAAIQFDAGGVP 187
Db 238 VIAIGDPLGKDLISRTVTQIGVSDVTVSMS---TSAGETISINVITQDAINFGSGGPL 294
QY 188 VNLGQVGVGNT--AASDNFQISQGGQGFALPIQGMVAIQIRSGGSGSFTVHIGTAPL 245
Db 295 LNTDGRVINGSMKISDDVE---GIGFAIFSDNDVKPIAEELLKSGQTEREYIG-VSML 349
QY 246 GLGVDDNN-----NGARVORVWGSAPASIGTGTGVITAVDCAPINSATA 292
Db 350 DLEQVQNTQEGTLGLFGSLQKNGVYIREVAGSGFAEKAGLKAEDIIIGLKREIDTUSE 409
QY 293 MADAL-NGHHPGDVTSV 308
Db 410 LRNLKYDKAKIGTVEV 426
RESULT 11
DEGP_ECOLI
ID DEGP_ECOLI STANDARD; PRT; 474 AA.
AC P09376; P15724;
DT 01-MAR-1989 (Rel. 10, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protease do precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR PTD OR B0161 OR Z0173 OR ECS0165.
OS Escherichia coli 0157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCB_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89057448; PubMed=3057437;
RA Lipinska B., Sharma S., Georgopoulos C.;
RT "Sequence analysis and regulation of the htra gene of Escherichia
RT coli: a sigma 32-independent mechanism of heat-inducible
RT transcription."
RL Nucleic Acids Res. 16:10053-10067(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Kura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 2.4-4.1 min (110,917-193,643 bp) region."
RL Nucleic Acids Res. 22:1637-1639(1994).
RN [3]
RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
RN [7]
RP SEQUENCE OF 1-50 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90323597; PubMed=2165018;
RA Quirk S., Bhatnagar S.K., Bessan M.J.;
RT "Primary structure of the deoxyguanosine triphosphate
RT triphosphohydrolase-encoding gene (dgt) of Escherichia coli."
RL Gene 89:13-18(1990).
RN [8]
RP SEQUENCE OF 1-16 FROM N.A.
RC MEDLINE=90207273; PubMed=2157212;
RA Wurgler S.M., Richardson C.C.;
RT "Structure and regulation of the gene for dGTP triphosphohydrolase
RT from Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 87:2740-2744(1990).
RN [9]
RP SEQUENCE OF 27-39, AND CHARACTERIZATION.
RX MEDLINE=90202693; PubMed=2180903;
RA Lipinska B., Zylicz M., Georgopoulos C.;
RT "The htra (DegP) protein, essential for Escherichia coli survival at
RT high temperatures, is an endopeptidase."
RL J. Bacteriol. 172:1791-1797(1990).
RN [10]
RP SEQUENCE OF 27-30.
RC STRAIN=K12 / W3110;
RX MEDLINE=98263247; PubMed=9600841;
RA Wilkins M.R., Gasteiger E., Tonella L., Ou K., Tyler M.,
RA Sanchez J.-C., Goolley A.A., Walsh B.J., Bairoch A., Appel R.D.,
RA Williams K.L., Hochstrasser D.F.;
RT "Protein identification with N and C-terminal sequence tags in
RT proteome projects."
RL J. Mol. Biol. 278:599-608(1998).
RN [11]
RP IDENTITY OF HTRA AND PROTEASE DO.
RX MEDLINE=91222240; PubMed=2025286;

Seol J.H., Woo S.K., Jung E.M., Yoo S.J., Lee C.S., Kim K.J.,
 Tanaka K., Ichihara A., Ha D.B., Chung C.H.,
 "Protease Do is essential for survival of *Escherichia coli* at high
 temperatures: its identity with the htra gene product.",
 Biochem. Biophys. Res. Commun. 176:730-736(1991).
 [12]
 RA DISULFIDE BOND.
 RX MEDLINE=22760311; PubMed=12878036;
 RA Skorko-Glonek J., Zurawa D., Tanfani F., Scire A., Wawrzynow A.,
 Narkiewicz J., Bertoli E., Lipinska B.;
 "The N-terminal region of HtrA heat shock protease from *Escherichia*
 coli is essential for stabilization of HtrA primary structure and
 maintaining of its oligomeric structure".
 Biochim. Biophys. Acta 1649:171-182(2003).
 CC -!- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
 CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IT CAN DEGRADE
 CC ICIA, ADA, CASEIN AND GLOBIN. SHARED SPECIFICITY WITH DEGO.
 CC -!- SUBUNIT: Multimeric.
 CC -!- SUBCELLULAR LOCATION: Periplasmic.
 CC -!- INDUCTION: By heat shock.
 CC -!- MISCELLANEOUS: HTRA IS INDISPENSABLE FOR BACTERIAL SURVIVAL AT
 CC TEMPERATURES ABOVE 42 DEGREES CELSIUS.
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M36536; AAA23994.1; -;
 CC EMBL; X12457; CAA30997.1; -;
 CC EMBL; D26562; BAB96738.1; -;
 CC EMBL; AE000125; AAC73272.1; -;
 CC EMBL; U70214; BAB08591.1; -;
 CC EMBL; AE005192; AAG54465.1; -;
 CC EMBL; AP002550; BAB33588.1; -;
 CC EMBL; M29955; AAA23717.1; -;
 CC EMBL; M31772; AAA23680.1; -;
 CC PIR; E85500; E85500.
 CC PIR; E90649; E90649.
 CC PIR; S45229; S45229.
 CC PDB; 1KI9; 03-APR-02.
 CC MEROPS; S01.273; -;
 CC SWISS-2DPAGE; P09376; COLI.
 CC EcoGene; EG10463; degP.
 CC InterPro; IPR009003; Cys_Ser_trypsin.
 CC InterPro; IPR001478; PDZ
 CC InterPro; IPR001254; Peptidase_S1.
 CC InterPro; IPR001940; Peptidase_SIC.
 CC Pfam; PF00595; PDZ; 2.
 CC Pfam; PF00089; trypsin; 1.
 CC PRINTS; PR00834; PROTEASES2C.
 CC SMART; SM00228; PDZ; 2.
 CC PROSITE; PS01061; PDZ; 2.
 CC Hydrolase; Serine protease; Heat shock; Periplasmic; Repeat; Signal;
 KW Complete proteome; 3D-structure.
 FT SIGNAL 1 26
 FT CHAIN 27 474
 FT DOMAIN 280 371
 FT PDZ 1.
 FT PDZ 2.
 FT ACT_SITE 131 131
 FT ACT_SITE 161 161
 FT ACT_SITE 236 236
 FT ACT_SITE 83 95
 FT DISULFID 10 10
 FT CONFLICT 46 46
 FT CONFLICT 102 102
 FT CONFLICT 192 192
 FT CONFLICT 467 474
 FT STVLQW -> RLPLPNVLSNPLFKTGSGPYNL (IN
 FT REV. 1).
 SQ SEQUENCE 474 AA; 49354 MW; 5482E596F74B6D5F CRC64;
 Query Match 17.2%; Score 291.5; DB 1; Length 474;
 Best Local Similarity 33.2%; Pred. No. 6e-10;
 Matches 88; Conservative 43; Mismatches 81; Indels 53; Gaps 11;
 QY 54 GTGIVIDPN-GVLTNNHVIAGATDINAFVSGSQTYGVGVDRVDTQDVAVLQRCAGG 112
 DB 114 GSGVILADKGYVYNNHVDNATVIKY-QLSDGRKFDAMKMGKDPKSDIALIGIQLPKN 172
 QY 113 LPS--AAIGGGVAVGEPVWAMNGSGGCGTPRAYGVRVAGLQTVQASDSLSLTGAE-ETLN 169
 DB 173 LTAIKMADSDALRVGDYTVVAINPFGLET--VTSGIVSALGR-----SGLNAENYE 222
 QY 170 GLIQFDAAIQFDAGGPPVNGLGQVGVGNVTA--ASDNFQLSQGGQGFAP-----IGQA 221
 DB 223 NFQITDAAINRNGSGGALVNLGELIGINTAILAPDGGNI---GIGFAIPSNVYKNTSQ 279
 QY 222 MALAGQIRSGGSPVTHIGPTAFGLGLGVVDNNGN-----GARQVRVVGSA 268
 DB 280 MVEYGVQVRG-----ELGIMGTLSLAKAMKVDQAQRGAQVSVQLPNSA 325
 QY 269 ASLGISTGDTITAVDGPINSATAM 293
 DB 326 AKAGIKAGDVITSINGKPISSFAAL 350
 RESULT 12
 ID HTOA HAEIN STANDARD; PRT; 466 AA.
 AC P45129;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable periplasmic serine protease do/hhoA-like precursor
 DE (EC 3.4.21.-).
 GN H11259.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd.";
 RL Science 269:496-512(1995).
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC -!- SIMILARITY: SEEMS TO BE AN INTERMEDIATE FORMS BETWEEN E.COLI HTRA
 CC (PROTEASE DO) AND HHOA.
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U32805; AAC22906.1; -;
 CC PIR; A64113; A64113.

```

DR MEROPS; S01.274; -.
DR TIGR; H1259; -.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
DR Hydroxylase; Serine protease; Periplasmic; Repeat; Signal;
KW Complete proteome.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 466 DO/HOAA-LIKE.
FT DOMAIN 270 361 PDZ 1.
FT DOMAIN 367 458 PDZ 2.
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 150 150 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 226 226 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 466 AA; 49434 MW; ED050A00047B5851 CRC64;

Query Match 17.0%; Score 289; DB 1; Length 466;
Best Local Similarity 32.6%; Pred. No. 8.2e-10;
Matches 93; Conservative 43; Mismatches 101; Indels 48; Gaps 11;

QY 49 NAVGAGTGIVDPN-QVLTNNHVIAGTDINAFVSGQTYGVVDVDRDQDVAVLQL 107
DB 98 NFRGLSGVVIINASKGYVLTNNHVIDGADKITV-QLQDGRFKAKLVGDEQSDIALVQL 156
QY 108 RGAGGLPSA--AIGGGVAVGEPVWAMGNSGGGGTTPRAVPGRVVALGQTVQASDSLTGAE 165
DB 157 EKPNSLTKFKPADSKLRVGDFTVAIGNFGLGQT--VTSGIVSALGRS-----TGS 207
QY 166 E-TLNGLIQDAAIQDAGGVPVNGLGQVGNMTAASDNFQLSGGGG--FAIPGQAM 222
DB 208 SGTIENYIQTDAAVNRGSGGALVNLGELIGINTAI---ISPSGNGAGIAFAIPSNQAS 264
QY 223 ATA-----GQIRSGGGSTVHIGTAFGLGLGVNDNGN-----GARVQVVS 265
DB 265 NLVQQLFEGQVRRG-----LLGKGGELNADLAKAFNVSAQGGAFVSEVLPK 312
QY 266 APAASLGISTGVITAVDGAPINSATAMADALNGHHFGSDVSVTW 310
DB 313 SAAEKAGLRAGDIITAMNQKISSFAIRAKIATTGAGKEISITY 357

RESULT 13
ID DEGP BARHE STANDARD; PRT; 503 AA.
AC P54925;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable periplasmic serine protease DO-like precursor (EC 3.4.21.-)
DE (Antigen Htra).
GN DEGP OR HTRA.
OS Bartonella henselae (Rochalimaea henselae).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bartonellaceae; Bartonella.
OX NCBI_TaxID=38323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 49882 / Houston 1;
RX MEDLINE=94299828; PubMed=8027347;
RA Anderson B., Sims K., Regnery R., Robinson L., Schmidt M.J.,
RA Goral S., Hager C., Edwards K.;
RT "Detection of Rochalimaea henselae DNA in specimens from cat scratch
RT disease patients by PCR."
RL J. Clin. Microbiol. 32:942-948(1994).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.

```

```

CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L20127; AAA97430.1; -.
DR MEROPS; S01.273; -.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
KW Hydroxylase; Serine protease; Periplasmic; Repeat; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 503 PROBABLE PERIPLASMIC SERINE PROTEASE DO-
FT DOMAIN 286 357 PDZ 1.
FT DOMAIN 419 466 PDZ 2.
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 503 AA; 54114 MW; 6CD9F4743282AF9E CRC64;

Query Match 16.8%; Score 285.5; DB 1; Length 503;
Best Local Similarity 27.4%; Pred. No. 1.4e-09;
Matches 107; Conservative 43; Mismatches 125; Indels 115; Gaps 15;

QY 12 LSQDRFADFALPLDPSAMVAQVGPVQVNTKLG----- 46
DB 47 MQQQGFAD-----IVSQVKPAVVSVQVSKNKKKEWFFSDFTSTPGFDQLPDQHP 96
QY 47 -----YN-----NAVAGTGVIVDPNGVLTNNHVIAGTDINAFSV 83
DB 97 LKFFQDFYNRDKPSNKSLSQRHLRPIAFSGFFISSDGIYVTTNNHVISDGTSY-AVVL 155
QY 84 GSGQTYGVVDVVDYDRTQDVAVLQRLRGAGGLPSAAIG--GGVAVGEPVWAMGNSGGGGTTP 141
DB 156 DGTGLNAKLIGTPTDRLAVLKNEXKFSYVDVFGDSSKLVRGWDVVAIGNPFLGGTV 215
QY 142 RAVPGRVVALGQTVQASDSLTGABETLNGLIQDAAIQDAGGVPVNGLGQVYGMNTAA 201
DB 216 TA--GIVSARGR-----DIGTG---VYDDFIQIDAAVNRGNSGGFTFDLNGKVYGVNTAI 265
QY 202 SDNFQLSGGGG--FAIPGQAMAIAGQIRSGGGSPVHIGTAFGLGVVDNNGGARV 259
DB 266 ---FSPGGNVGIAFAIPAATANEVQQL-----IEKGLVQRWLGVQI 306
QY 260 QRV-----VG-----SAPASLSGISTGEVITAVDGAIPNSATAMADALNG 299
DB 307 QPVTKEITSDSISGLKEAKGALITDPLKGPAAKAGIKAGDVIISVNGEKINDVRDLAKRIAN 366
QY 300 HHPGDVSV-TWQTKSGGTRTGNVTLAEGP 328
DB 367 MSPGETVTLGW--KSGKEENIKVKLDSMP 394

RESULT 14
ID DEGP CHLTR STANDARD; PRT; 497 AA.
AC P18584; O84830;
DT 01-NOV-1990 (Rel. 16, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-) (59 kDa)

```

```

114 114 PSNAITGG--VAVGPPVAMGNSGGCGTTPRVPGRVVALGQT---VQASDLSLTGAET 167
185 185 PFLTFGNSDQLGDMWAIAGNPFGLQAT--VTVGVISAKGRNLHIVDFED-----234
168 168 LNLGLIQDAAIOPGDAGGPPVAVGLGVGMGNTAASDNFQLSG--CGQFAIFIQOMAJA 225
235 235 ---FIOTDAAINPNSGGGPELLNNGVIGVNTAIVSG---SGYIGIGFAIPSLMAKRV 288
226 226 GQIRSGGSPFVHIQPTAFGLGIVVDN-----NNGGARVQVRVVGAPASLIGISIG 276
289 289 DQLISDQGVQTRGFLGVT---LQPIDSELAATCYKEKVGALVTDVVGKSPAKAGLROE 344
277 277 DVITAVDGPAINSATAMADALNGHPGDVVISVTWQTSKGGTRTGNVTLAEGP 328
345 345 DIVAVNGKEVSELSALRNAISLMMPGTRV-VLKIVREKGTIPIPTVTQIP 395

RESULT 15
DEGP_BUCAI STANDARD; PRT; 478 AA.
AC P5732;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR BU228.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiote bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_taxid=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RC MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -! SIMILARITY: Belongs to peptidase family S2C.
CC -! SIMILARITY: Contains 2 PDZ/DHR domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC
CC EMBL; AP001118; BAB12943.1; --
CC MEROPS; S01.273; --
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ; 2.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SM00228; PDZ; 2.
CC PROSITE; PS50106; PDZ; 1.
CC Hydrolase; Serine protease; Repeat; Signal; Complete proteome.
CC SIGNAL 1 26 POTENTIAL,
CC CHAIN 27 478 PROBABLE SERINE PROTEASE DO-LIKE.
CC DOMAIN 116 254 CATALYTIC.
CC DOMAIN 281 372 PDZ 1.
CC DOMAIN 397 469 PDZ 2.
CC ACT_SITE 133 133 CHARGE RELAY SYSTEM (POTENTIAL).
CC ACT_SITE 163 163 CHARGE RELAY SYSTEM (POTENTIAL).
CC ACT_SITE 238 238 CHARGE RELAY SYSTEM (POTENTIAL).
CC DISULFID 57 99 BY SIMILARITY.
CC SEQUENCE 478 AA; 52230 MW; 868E8732CAC50629 CRC64;
SQ

```

```
Query Match      16.7%; Score 284; DB 1; Length 478;
Best Local Similarity 30.6%; Pred. No. 1.6e-09;
Matches 88; Conservative 51; Mismatches 109; Indels 40; Gaps 12;

QY      54 GTGIVIDPN-GVVLNNHVIAGATDINAFSVSGGOTYGVVDVGYDRTOADVAVLQLRGAGG 112
Db      116 GSGVIINADKGYAVTNNHVVENANKIQV-QLSDGRRYEARVIGKDSRDIALIQLKNANN 174

QY      113 LPSAAI--GGGVAVGEFVVVAMGNSGGQGTTPRAVGRVVVALGQTVQASDSLTGAE-ETLN 169
Db      175 LSEIKIADSDNLRVGDTVAIGNPYGLGET--VTSGIISALGR-----SGLNIEHYE 224

QY      170 GLIQFDAATOPGDAGPVGNGI-GVVGMNTA--ASDNFOLSOGGOGFAIPIQAMAIAGQ 227
Db      225 NFIQTDAALNRGNSGGALVNLKGLIGINTAILAPDGGNI---GIGFALPCNMVKNLTAQ 281

QY      228 IRSGGGSPTVHIGPTAFLGLGVVD-----NNGNGARVQVRVVGSAAPAASLGISTGDVI 279
Db      282 MVQFG--QVRRGELGIMGMELNSDLAQIMKINSQKGFVSRVLPNSSAFEAGIKAGDII 338

QY      280 TAVDGNPINSATAMADALNGHHPGDVISVTWTKSGGTRTG---NVTL 324
Db      339 ISLNRKPISFSLSRAEIGS-----LPVATKVELGVFREGRIKNITY 380
```

Search completed: June 22, 2004, 17:16:48
Job time : 6.58295 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:55:54 ; Search time 26.9826 Seconds

(without alignments)
3858.816 Million cell updates/sec

Title: US-09-886-349A-6

Perfect score: 1698

Sequence: 1 MHHHHHAPPALSQDRFADF.....QTKSGGTRGNVTLAEGPPA 330

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.25.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_proteus.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacterioph.*

17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1642	96.7	355	16 O07175	O07175 mycobacteri
2	1642	96.7	355	16 Q7U2S9	Q7U2S9 mycobacteri
3	1228.5	72.3	361	2 Q50320	Q50320 mycobacteri
4	1195	70.4	354	16 Q9CCY9	Q9CCY9 mycobacteri
5	457.5	26.9	446	16 Q8VKA4	Q8VKA4 mycobacteri
6	457.5	26.9	464	16 Q7U0X2	Q7U0X2 mycobacteri
7	456.5	26.9	464	16 Q53896	Q53896 mycobacteri
8	447	26.3	382	16 Q9CD67	Q9CD67 mycobacteri
9	447	26.3	452	2 Q9Z5G6	Q9Z5G6 mycobacteri
10	394.5	23.2	519	16 Q93J30	Q93J30 streptomyc
11	393	23.1	375	16 Q8DC87	Q8DC87 streptomyc
12	390.5	22.4	542	16 Q9FBK9	Q9FBK9 streptomyc
13	381	22.0	472	16 Q82FM9	Q82FM9 streptomyc
14	381	22.4	619	16 Q82IL8	Q82IL8 streptomyc
15	376	22.1	473	16 Q8PR17	Q8PR17 streptomyc
16	371	21.8	394	16 P72780	P72780 streptomyc

ALIGNMENTS

RESULT 1

O07175	O07175	PRELIMINARY;	PRT;	355 AA.
AC	O07175;			
DT	01-JUL-1997 (Tremblrel. 04, Created)			
DT	01-JUL-1997 (Tremblrel. 04, Last sequence update)			
DT	01-OCT-2003 (Tremblrel. 25, Last annotation update)			
DE	Hypothetical protein (Serine protease, putative).			
OS	PEPA OR RV0125 OR MTC1418B.07 OR MT0133.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RC	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544(1998).			
RN	[2]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,			
RA	Decher A., Utterback I., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W.;			
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			

O8ytf9 anabaena sp
Q8dmv9 synchococc
Q8gdt3 bifidobacte
Q8na10 corynebacte
Q7vsc8 prochloroco
Q89j15 clostridium
Q97gd5 clostridium
Q8r756 thermoanaer
P73354 synchococc
Q89rp2 bradyrhizob
Q8pmv4 xanthomonas
Q9kjne myxococcus
Q7u495 synchococc
Q8xpt5 ralstonia s
Q8dl28 synchococc
Q31388 bradyrhizob
Q89gj8 bradyrhizob
Q7u8k9 synchococc
Q87c10 xylella fas
Q9ba3 xylella fas
Q7u604 synchococc
Q9s2r5 streptomyc
Q92qe6 rhizobium m
Q8rtk2 xanthomonas
Q8pb56 xanthomonas
Q8yvh0 anabaena sp
Q83752 treponema p
Q9rtk4 deiococcus
Q7va24 prochloroco

```

CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: Z96071; CAB09453.1; -.
DR RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
DR RT "The complete genome sequence of Mycobacterium bovis.";
DR RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR DR EMBL: BX248334; CAD92991.1; -.
DR KW Hydrolase; Complete proteome.
DR SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 96.7%; Score 1642; DB 16; Length 355;
Best Local Similarity 99.7%; Pred. No. 7.4e-81;
Matches 322; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPQVNVNTKLGYNNAVAGTGVIDPENGVLVT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPQVNVNTKLGYNNAVAGTGVIDPENGVLVT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAALGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAALGGVAVGEP 152
QY 128 VVAMNGSGGGGTPRAVPGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPV 187
DB 153 VVAMNGSGGGGTPRAVPGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPV 212
QY 188 VNGLGQVVMNTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPTVHIGPTAFGL 247
DB 213 VVAMNGSGGGGTPRAVPGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPV 272
QY 248 GVVDNNGGARVQVWVGSAPASLIGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 307
DB 273 GVVDNNGGARVQVWVGSAPASLIGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VTWQTKSGGTRTGNVTLAEGPPA 355

RESULT 3
Q50320 PRELIMINARY; PRT; 361 AA.
AC Q50320;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 34KDa protein precursor.
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JDS8/107;
RX MEDLINE=95005449; PubMed=7921248;
RA Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
RT "Identification and characterisation of a putative serine protease
expressed in vivo by Mycobacterium avium subsp paratuberculosis.";
RL Microbiology 140:1977-1982(1994).
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR EMBL: Z23092; CAB0638.1; -.
DR FIR; S47170; S47170.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ_S1.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.

```

DR	Pfam; PF00595; PDZ; 1.
DR	Pfam; PF00089; trypsin; 1.
DR	PRINTS; PRO0834; PROTEASES2C.
DR	PRINTS; PRO0839; V8PROTEASE.
DR	SMART; SM00228; PDZ; 1.
DR	PROSITE; PS50106; PDZ; 1.
KW	Hydrolase; Protease; Serine protease; Complete proteome.
SQ	SEQUENCE 354 AA; 35265 MW; 612F23261BC9EA4A CRC64;
Query Match 70.4%; Score 1195; DB 16; Length 354;	
Best Local Similarity 71.7%; Pred. No. 7.8e-57;	
Matches 231; Conservative 41; Mismatches 46; Indels 4; Gaps 2;	
QY	9 PPALQSQRADPADPAIPPLDPSAMVAQVGPVVNINTKLGYNNAVGAGTGIVIDNGVVLTN 68
DB	37 PSTLALDRESNPPLFLNPAAWA--POVVALSTELGYN SAVGAGTGVIDSSGVVLTN 93
QY	69 NHEVIAGATDINAFSGSGOTGYDVVVDYDRTQDVAVQLRQGAGLPSSAAIGGGVAVGEVP 128
DB	94 NHVISGATDISAFDYGNGKTYGDVVVDYDRTQDVAVQLRGSNLFTAVIGGDVAIGEPI 155
QY	129 VMANGNSGGCGGTTRAVPRGVVALGQTVOASDSL TGAEETLNGLIQFDALIQDGAGPVP 188
DB	154 VALGNTTGGGGPLSPVLPGRVWALNQTVQASEPLTGAQETLSGLIQVDAPIKPGDSGGPVP 213
QY	189 NLGLGVVMGMTAASNDFQLSQGGGF AIPFGQAMA IAGAIRSGGSGPTVHIGTAFGLG 248
DB	214 NSRGQVVMGMTAATDN YKM-LGGQGF AIPFGQAMEVVGAI RSGAGSN TVHIGTAFGLG 272
QY	249 VVDNRNGNGARVORVVGSAFAASLGISTGVITTA VD GAPINS ATAMADALNHGHFDPVISV 308
DB	273 VLDDNNGNGARVARV VATGPAAMAGISVGDIIITSVD GPPISEATAMTNLVLPHPGETVAV 332
QY	309 TWQTKSGGTRTCGNVT LAEGPPA 330
DB	333 NRSAGGGDLTANVT LAEGPPA 354
RESULT 5	
Q8VK44 PRELIMINARY; PRT; 446 AA.	
ID	Q8VK44
AC	01-NAR-2002 (TrEMBLrel. 20, Created)
DT	01-NAR-2002 (TrEMBLrel. 20, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Heat shock protein HtrA, putative.
GN	Mt1011
OS	Mycobacterium tuberculosis.
OC	Bacteria; Actinobacteria; Actinomycetales;
CC	Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX	NCBI_Taxid=1773;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CDC 1551 / Oshkosh;
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA	Peterson J.F., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA	Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA	Bissh W.;
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and
RT	laboratory strains";
RL	Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; A5006985; AAK45259.1; -.
DR	TIGR; MT1011; -.
DR	GO; GO:0004295; F:trypsin activity; IEA.
DR	GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR	InterPro; IPR009003; Cys_Ser_trypsin.
DR	InterPro; IPR001478; PDZ; 1.
DR	InterPro; IPR001254; Peptidase_S1.
DR	InterPro; IPR001940; Peptidase_S1c.
DR	Pfam; PF00595; PDZ; 1.
DR	Pfam; PF00089; trypsin; 1.

```

DR PRINTS; PRO0834; PROTEASES2C.
DR SMART; SMO0228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
SQ SEQUENCE 446 AA; 44484 MW; 54170CBEA8FE72B CRC64;

Query Match
Best Local Similarity 26.9%; Score 457.5; DB 16; Length 446;
Matches 121; Conservative 56; Mismatches 118; Indels 31; Gaps 10;

QY 21 PALPLDPSA---MVAQGVGVVNTKLGYNNAVAGTGTIVDPNGVLTNNHVIAGAT- 76
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 129 PAANMPGSGVEQVAAKVPSVVMLETLGRQSE--EGSGIILSAEGLIITNNHVAATAK 186
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 77 -----DINAFVSGQTYGVGVYDRTPQDVAVLQLRGAGLPSAAIGGG--VAVGEP 127
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 187 PPLGSPPKTTVTFSDGRTAPFTVVGADPTSDIAVVRVGVVGLTPIISGSSDLRVGP 246
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 128 VVAMGNSGGCGTTPRAVGRVVALGQTVQASDSLTGABETLNGLIQFDAAIQPDAGGPV 187
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 247 VLAIGSPGLGEGT--VTGTIVSALNRPVSTTGE--AGNQNTVLDAIQTDAAINPGNSGGAL 303
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 188 VNLGLGVVGNMTA---ASDNFQLSQG--GCGFAIPIGQAVAIAGQIRSGGSPVTHIGP 241
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 304 VVNAQLGVNSAIATLGADSDAQSGISGLGFAIPVDQAKRIADELISGTGA-----S 357
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 242 TAFILGLGV--DNNNGGARVORVVGSAAPASGLISTGDTAVDVGAPINSATAMADALNGH 300
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 358 HASLGVQVNTKDTTPGAKIVVAVAGGAANAGVPKGVVTVKDDRPINSADALVAARSK 417
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 301 HPGDVISVTWQTKSGGTRTGNVTIAE 326
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 418 APGATVALTFODPGSGGRTVQVTLGK 443
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 6
QYU0X2 PRELIMINARY; PRT; 464 AA.
AC QYU0X2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable serine protease (Serine proteinase) (EC 3.4.21.-).
GN MB1009.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AP2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garner T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duchoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248337; CAD93870.1;
DR Hydrolyase; Complete proteome.
SQ SEQUENCE 464 AA; 46436 MW; AE93A4BB3FFA9BE3 CRC64;

Query Match
Best Local Similarity 26.9%; Score 457.5; DB 16; Length 446;
Matches 121; Conservative 56; Mismatches 118; Indels 31; Gaps 10;

QY 21 PALPLDPSA---MVAQGVGVVNTKLGYNNAVAGTGTIVDPNGVLTNNHVIAGAT- 76
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 147 PAANMPGSGVEQVAAKVPSVVMLETLGRQSE--EGSGIILSAEGLIITNNHVAATAK 204
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 77 -----DINAFVSGQTYGVGVYDRTPQDVAVLQLRGAGLPSAAIGGG--VAVGEP 127
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 205 PPLGSPPKTTVTFSDGRTAPFTVVGADPTSDIAVVRVGVVGLTPIISGSSDLRVGP 264
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

PRINTS; PRO0834; PROTEASES2C.
SMART; SMO0228; PDZ; 1.
PROSITE; PS0106; PDZ; 1.
SEQUENCE 446 AA; 44484 MW; 54170CBEA8FE72B CRC64;

Query Match
Best Local Similarity 26.9%; Score 457.5; DB 16; Length 446;
Matches 121; Conservative 56; Mismatches 118; Indels 31; Gaps 10;

QY 21 PALPLDPSA---MVAQGVGVVNTKLGYNNAVAGTGTIVDPNGVLTNNHVIAGAT- 76
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 147 PAANMPGSGVEQVAAKVPSVVMLETLGRQSE--EGSGIILSAEGLIITNNHVIAGAT- 76
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 77 -----DINAFVSGQTYGVGVYDRTPQDVAVLQLRGAGLPSAAIGGG--VAVGEP 127
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 205 PPLGSPPKTTVTFSDGRTAPFTVVGADPTSDIAVVRVGVVGLTPIISGSSDLRVGP 264
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

PRINTS; PRO0834; PROTEASES2C.
SMART; SMO0228; PDZ; 1.
PROSITE; PS0106; PDZ; 1.
SEQUENCE 446 AA; 44484 MW; 54170CBEA8FE72B CRC64;

Query Match
Best Local Similarity 26.9%; Score 457.5; DB 16; Length 446;
Matches 121; Conservative 56; Mismatches 118; Indels 31; Gaps 10;

QY 21 PALPLDPSA---MVAQGVGVVNTKLGYNNAVAGTGTIVDPNGVLTNNHVIAGAT- 76
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 147 PAANMPGSGVEQVAAKVPSVVMLETLGRQSE--EGSGIILSAEGLIITNNHVIAGAT- 76
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 77 -----DINAFVSGQTYGVGVYDRTPQDVAVLQLRGAGLPSAAIGGG--VAVGEP 127
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 205 PPLGSPPKTTVTFSDGRTAPFTVVGADPTSDIAVVRVGVVGLTPIISGSSDLRVGP 264
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```


QY 18 ADFPALPDSAMVAQVG-----PQVNNITKLGNNNAVAGAGTGIVIDPNGVLTNNHVA 73
 DB 129 ASVPAANN-PSGSVEQVAVKVFESVLMLETLDRQSE--EGSGVILSADGLILTNHVA 185
 QY 74 GA-----TDINAFSGSQTVGVVDVQVDRDQVAVLQRLGAGGLPSAAIG 119
 DB 186 VAAKPGGGPGGLSPKTVTF--DRTASFTVWGADPTSDIAVVRVQSGISLTITWG 242
 QY 120 GG--VAYGEPVAMNGSGGGGTPRAVPGRVVALGQTVQASDLSLTGABETLNGLIQFDDA 177
 DB 243 SSADLRVQGVAVGSLPLAGT--VTSIGVSLNRPVSTTGB-SGNQNTVLDAIQTDAA 299
 QY 178 IQPGDAGPVPVNGLGQVVGNTAA-----SDNFOLSQGGGQGAIPGAMATAGQIRSG 231
 DB 300 INFNSGGALVNGGQLGVNSAIALTGADSGDAQSGSIGLGFAPVDAQRIADELIST 359
 QY 232 GGSFTVHIGTAFGLGLGWDNNGN-GARVQVVGSGAPAAALGISTGIVITAVDGPAINSA 290
 DB 360 G--KATH-----ASLGQVATDKCTPGAKVVDVAVAGGAANAAPKGVLTAKVDRLLISA 413
 QY 291 TAVADALNHGHPGVISVTVWQTSKGGTRTGNVTIAE 326
 DB 414 DALVAAVRSKAPGDKVSLTYQDSGSSRTVQVTLGK 449

RESULT 10
 Q93J30 PRELIMINARY; PRT; 519 AA.
 AC Q93J30;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative protease;
 GN SC03977 OR SCBAC253.14.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Collins M., Harris D.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Batenan A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RL "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RN Nature 417:141-147(2002).

DR EMBL; AL939118; CAC44701.1; --
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_ser_trypsin.
 DR InterPro; IPR001478; PDZ_Src.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR SMART; SM00020; Tryp_Src; 1.
 DR KW Hydrolyase; Protease; Serine protease; Complete proteome.
 SQ SEQUENCE 519 AA; 50327 MW; SCB8D5F0CC19E428 CRC64;

Query Match 23.2%; Score 394.5; DB 16; Length 519;
 Best Local Similarity 34.1%; Pred. No. 1.3e-13;
 Matches 108; Conservative 52; Mismatches 128; Indels 29; Gaps 11;

QY 28 SAMVAQVGPVNNITKLGNNNAVAGAGTGIVIDPNGVLTNNHVAIDINAEVS--GS 85
 DB 205 AGVAAKALPSTVITQAE--GSNCGEGTGTGFVFDKEGHIVTNNHVAEAVDGGKLSATFPN 263
 QY 86 GQTVGVVDVVDRTQDVAVLQRLGA--GGLPSAAIAGG--VAVGEPVAMNGSGGGGTTPR 142
 DB 264 GKDYDAEVVGHAGQYDVAIVIKLENAPSDLKELALGDSKDVAVGDSITAIAGAPGLSNT-- 321
 QY 143 AVPGRVVALGQTVQASDLSLTGABETLNGLIQFDDAIOFGDAGGVPVNGLGQVGNMTA-- 200
 DB 322 VTTGIISAKNRPVASSDGSADSKASYMSALQTDASINFGNSGGLPLDAGQNVIGINSAIQ 381
 QY 201 --ASDNFOLSQGGQ---GFAIPICOMAIAGQISSGGSPVHIGTAFGLGLGVVDNNGN 255
 DB 382 STNGGGTGTGAGSIGLGFAPVNAQFAQLKSKGPVYAKIGASVSL-----ETTNN 436
 QY 256 GARV--QRVVGSG-----APAAISLGISTGIVITAVDGPAINSATAMADALNHGHPGVIS 307
 DB 437 GAKLTEQGVGSGSDPVEKGGPADDAKLPQDVITKLDNRVIDSGPTLIGEITWTHKPGDEV 496
 QY 308 VTWQTSKGGTGTGNVTIL 324
 DB 497 VTYE--RGCKQHTAEVTL 512

RESULT 11
 Q93J30 PRELIMINARY; PRT; 375 AA.
 AC Q93J30;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Serine protease.
 GN TLR2436.
 OS Synchococcus elongatus (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.
 OX NCBI_TaxID=32046;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1;
 RX MEDLINE=22225144; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Kato H., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
 RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the thermophilic cyanobacterium
 Thermosynechococcus elongatus BP-1";
 RL DNA Res. 9:123-130(2002).
 DR EMBL; AP005377; BAC0988.1; --
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.

DR GO: 0006508; P: proteolysis and peptidolysis; IEA.
DR InterPro: IPR009003; Cys_Ser_trypsin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR001940; Peptidase_S1C.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS50106; PDZ; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 375 AA; 39724 MW; 0D2602EB534915F5 CRC64;

Query Match 23.1%; Score 393; DB 16; Length 375;
Best Local Similarity 33.6%; Pred. No. 1e-13;
Matches 108; Conservative 50; Mismatches 113; Indels 50; Gaps 10;
RY 10 PALSQDRPADFPALPLDPSAMVAQVGPQVNVNTKLGYNNAVAGAGTGIVDPNGVLTN 68
Db 67 PLSPPFPFPFPFGALPQE-----DLRGQSGFIDPSGIWVN 108
QY 69 NHVIAAGATDINAFSVSGGTGVVGVYDRTQDVAVLQRLG-AGGLPSAAIGGG--VAVG 125
Db 109 AHVVSQADTAVV-RLKDGVRFEVGRGVDEVSALAIVKLKGVTEPLTAPLGDSSEVKVG 167
QY 126 EPVWANGSGGGGTTPRAVPGRVVVG--QTVAQSDSLTGAEETLNGLIQDAAIQGDA 183
Db 168 DWAIAGNPLGDNT-----VTGIISTLHRSAAQVGPDKLDFIQTDAAINPNS 219
QY 184 GGPVWNGLGQVVMNTAASDNFQSGGQGFPAIPQIQAAMAIAGQIRSGGSGTWHIG--P 241
Db 220 GGPLLNEAGEVIGINTA----IRADAMGIGFPAIPKAKALQARLIRGEKIQHAVIGIQM 275
QY 242 TAFGLGVVDNNG-----NGARVQVVGSGAPASLSIGSTGDIVTAVDGAPINSA 290
Db 276 TTTFTPAKAKENANPNPVLPEVNGVLVQLVNTPAKAGLRGWDVITAVDGPITSA 335
QY 291 TAVADALNGHFGDVISVTWQ 311
Db 336 DQLQITVDSAAVGVNLTVQ 356

RESULT 12

Q9FBK9 PRELIMINARY; PRT; 542 AA.
ID Q9FBK9
AC Q9FBK9
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative protease.
GN SC05149 OR SC98.12.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8943436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Gobie A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser H., Lark L., Murphree L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL939122; CAC01350.1; -
GO: GO:0008233; F: peptidase activity; IEA.
DR GO: GO:0004295; F: trypsin activity; IEA.
DR GO: GO:0007242; P: intracellular signaling cascade; IEA.
DR GO: GO:0006508; P: proteolysis and peptidolysis; IEA.
DR InterPro: IPR009003; Cys_Ser_trypsin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001254; Peptidase_S1.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS50106; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 542 AA; 53761 MW; 620F7D889DF7212D CRC64;
Query Match 23.0%; Score 390.5; DB 16; Length 542;
Best Local Similarity 32.3%; Pred. No. 2.2e-13;
Matches 111; Conservative 56; Mismatches 132; Indels 45; Gaps 11;
QY 9 PPALSQDRPADFPALPLDPSAMVAQVGPQVNVNTKLGYNNAVAGAGTGIVDPNGVLTN 68
Db 217 PEAAERD-----PDSVAGIAARALPSVVTLLH--VSGEAGTGTFVLDGRGHLTN 266
QY 69 NHVI--AGATDINAFSVSGGTGVVGVYDRTQDVAVLQRLGAGGLPSAAIGG--GVAV 124
Db 267 NHVVEPAGSGEITVTFTNSGDTAEAVVGRDSGYDLAVVKVGTGLTPLPLGNSDNVRV 326
QY 125 GEPVWANGSGGGGTTPRAVPGRVVVGALQTVQA--SDSLTGAEETLNGLIQDAAIQGDA 183
Db 327 GDPVVAIGAPFDLAGT--VTSGIISAKERPITAGGEGSGSDISYVDALQTDAPINPNS 384
QY 184 GGPVWNGLGQVVMNTA-----ASDNFQSQGGQGFPAIPQIQAAMAIAGQIRSGGGS 234
Db 385 GGPLLDARGAIGINSAIRSADSGSTESDDGQAGSIGLGFPAIPINQKRVAAEEL----- 438
QY 235 PTVHIGTAFGLGV-VDNN--NGARVQ-----RVVGSAPASLSIGSTGDIVTAVD 283
Db 439 --INTGKAHPVIGITLDNMTYTGDAEVSAKSGDGPVTTGGEGAKGKPGDIVTAVD 496
QY 284 GAPINSATAMADALNGHHPGDVISTVTWQTSKGGTRTGNVTIAEG 327
Db 497 GQRVHSGEELIVKTRAPRPGDRLELTLDQDKETKVSILVLSGG 540

RESULT 13

Q82FM9 PRELIMINARY; PRT; 472 AA.
ID Q82FM9
AC Q82FM9
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative protease.
GN SAV4223.
OS Streptomyces avermitilis.

01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
Putative serine proteinase.
SAV3115.
Streptomyces avermitilis.
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomyces.
SAV3115.

Streptomycineae; Streptomycetaceae; Streptomycetes.
NCBI_taxid=33903;
[1]
SEQUENCE FROM N.A.
STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
MEDLINE=21477403; PubMed=11572948;
Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
Shinohe M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
"Genome sequence of an industrial microorganism Streptomycetes
avermitilis: deducing the ability of producing secondary
metabolites.";
Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
[2]

omelitis"; deducing the ability of producing secondary metabolites.";
Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
[2]
SEQUENCE FROM N.A.
STRAIN=MA-4580 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
MEDLINE=22608306; PubMed=1269252;
Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
Sakaki Y., Hattori M., Omura S.;
"Complete genome sequence and comparative analysis of the industrial
microorganism *Streptomyces avermitilis*.";
Nat. Biotechnol. 21:526-531(2003).

"Complete genome sequence and comparative analysis of the industrial microorganism *S. pasteurii* strain *avermittalis*." Nat. Biotechnol. 21:526-531 (2003).

EMBL: AP005033; BAC70826.1; -

GO: GO:0004395; P:trypsin activity; TEA.

GO: GO:0007242; P:intracellular signaling cascade; TEA.

GO: GO:0006508; P:proteolysis and peptidolysis; TEA.

Interpro: IPR009003; Cvs Ser trypsin.

GO: GO:0004295; F-trypsin activity; IEA.
GO: GO:0007242; P:intracellular signaling cascade; IEA.
GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
InterPro: IPR009003; Cys Ser trypsin.
InterPro: IPR001478; PDZ.
InterPro: IPR001254; Peptidase S1.
InterPro: IPR001940; Peptidase_S1C.
Pfam: PF00335; PDZ; 1.
Pfam: PF00089; trypsin; 1.
PRINTS; PR00834; PROTEASES9C.
SMART; SM00228; PDZ; 1.
Complete proteome.
SEQUENCE 619 AA; 64251 MW; D2D3C72A57D3B3A6 CRC64;

PRINTS, F800634, F8018352C.
SMART; SM00228; PDZ; 1.
Complete proteome.
SEQUENCE 619 AA; 64251 MW; D2D3C72A67D3B3A6 CRC64;

Very Match	22.4%;	Score 381;	DB 16;	Length 619;
Local Similarity	33.7%;	Pred. No. 8.4e-13;		
Conservative	54;	Mismatches 126;	Indels 44;	Gaps 12

13 SQORFADFPALPLDPSAMVAQVGPQVVNINTKLYNNNAVGAAGTGVITDPNGVVLTNHHVI 72
|::||| | :
298 SEERAADSVA-----GIAARALPSPVTILHVK--GSAAEGTGTGFVLDGRGHILTNNHVV 349

```
. 73 --AGATDINAFSGSQTYGVVDWVGDRDTQDAVLQLRGAGLPSAAITGG--GVAVGEPV 128
      |||::: : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
350 EPAGSSGETSVTFESGETAKATWGRRUGSYDLAYVKVSGVGGLKMPILGNSDNVGVGDVP 409
```

129 VAMNGSGGGTTPRAVGRVVALGQTVA-SDSLTAEEETLNGLIQFDAAIQPDAGGPV 187
 410 VAIGAPFDIANT--VTSGITSAKERPTAGGKGGSDVSVDALORDAPINPGSGGPL 467

188 VNGLGVVGMNTAA----SDNPQLSQGQ--GPAIPIGQMAIAGIRSGGGSPTVHIG 240
:: :|||:: :|||:: :|||:: :|||:: :|||:: :|||:: :|||:: :|||:: :|||
468 LNSKAPVVGINSAFPSADSSDSOSGOAGSIGLGAIPVNCAKRAVEELINTGRATHPVIG 527

241 PTAFLGLGVDDNNGCARV-----QRVVGSAAPAAASLGISTGDIVTAVDGAAPSATA 292
:
: : | : |||
508 VMTNM-----DYVTNDCAIVGMVSNNGGSDYTPGAGBQNPAGICAGDVITFVDVGDTHSGRP 582

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:59:04 ; Search time 9.233 Seconds
(without alignments)
3438.018 Million cell updates/sec

Title: US-09-886-349A-6
Perfect score: 1693
Sequence: 1 MHHHHHAPPALSQDRFADF.....QTKSGGTRTGNVTLAEGPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR 78: **
2: PIR2: **
3: PIR3: **
4: PIR4: **

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1642	96.7	355	2 F70983	probable serine pr
2	1228.5	72.3	361	2 S47170	hypothetical prote
3	1195	70.4	354	2 A87242	probable secreted
4	456.5	26.9	464	2 C70821	probable serine pr
5	447	26.3	382	2 H86930	probable secreted
6	447	26.3	452	2 T45448	probable serine pr
7	371	21.8	394	2 S74643	proteinase hhoA (E
8	365	21.5	407	2 AG2150	serine proteinase
9	350	20.6	433	2 H97199	htrA-like serine p
10	349	20.6	452	2 S77538	serine proteinase
11	335	19.7	514	2 A82581	periplasmic protei
12	332.5	19.6	362	2 T35287	probable secreted
13	330	19.4	416	2 AB2057	serine proteinase
14	329	19.4	398	2 B71284	probable periplasm
15	329	19.4	441	2 E75357	probable periplasm
16	326.5	19.2	429	2 AD3894	serine proteinase
17	316.5	18.6	474	2 F83550	serine proteinase
18	316	18.6	355	1 JC6052	trypsin-like prote
19	316	18.6	355	2 D91142	trypsin-like prote
20	315	18.6	355	2 G85987	trypsin-like prote
21	314.5	18.5	401	2 AD2451	serine proteinase
22	312	18.4	408	2 H86891	exported serine pr
23	310.5	18.3	352	2 E82307	trypsin-like prote
24	309.5	18.2	455	2 AB0909	serine protease (E
25	309	18.2	416	2 S75445	proteinase hhoB (E
26	307.5	18.1	348	2 H96956	serine protease Do
27	307.5	18.1	455	2 C91142	serine endoprotein
28	306	18.0	455	2 F85987	serine endoprotein
29	305.5	18.0	513	2 AD3418	proteinase DO (EC

30 304.5 17.9 513 2 140060 serine proteinase
31 303.5 17.9 474 1 140059 htrA-like protein
32 303.5 17.9 474 2 AI3349 proteinase DO (EC
33 302.5 17.8 455 2 JC6051 trypsin-like prote
34 301 17.7 328 1 D69109 serine proteinase
35 299.5 17.6 356 2 AC0909 trypsin-like prote
36 298 17.6 459 2 F72359 periplasmic serine
37 296 17.4 457 2 AG0433 proteinase (EC 3.4
38 295 17.4 499 2 B81914 probable periplasm
39 293 17.3 475 1 S15337 heat shock protein
40 291.5 17.2 363 2 C87336 serine proteinase
41 291.5 17.2 449 2 A89643 serine proteinase
42 291.5 17.2 474 2 S45229 proteinase DO (EC
43 291.5 17.2 474 2 B85500 proteinase DO (EC
44 291.5 17.2 474 2 E90649 proteinase DO (EC
45 291 17.1 476 2 H71936 proteinase DO - He

ALIGNMENTS

RESULT 1

F70983
probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002
C:Accession: F70983
R:Coile, S.T.; Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome : A:Reference number: A70500; MUID:98295987; PMID:19634230
A:Accession: F70983
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Residues: 1-355 <COL>
A:Molecule type: DNA
A:Cross-references: GB:Z96071; GB:AL123456; NID:G3242254; PIDN:CAB09453.1; PID:G2181967
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: pepA
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp:

Query Match	96.7%	Score 1642;	DB 2;	Length 355;
Best Local Similarity	99.7%	Pred No. 7.8e-90;	Mismatches 1;	Indels 0; Gaps 0;
Matches 322;	Conservative 1;			
QY	8	APPALSDQRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGCTGVIDPENGVLIT 67		
Db	33	APPALSDQRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGCTGVIDPENGVLIT 92		
QY	68	NNHVIAGATDINASVSGSGQTYGVYDVYDRTQDVAVQLRGAGLPSAIGGSAVAYGEP 127		
Db	93	NNHVIAGATDINASVSGSGQTYGVYDVYDRTQDVAVQLRGAGLPSAIGGSAVAYGEP 152		
QY	128	VVMNGSGGGGTTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPDAGGPV 187		
Db	153	VVMNGSGGGGTTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPDAGGPV 212		
QY	188	VNGLGVVGMNTAASDNFQISQGGQFALPIGQAMATAGQIRSGGSGPTVHIGTAFILGL 247		
Db	213	VNGLGVVGMNTAASDNFQISQGGQFALPIGQAMATAGQIRSGGSGPTVHIGTAFILGL 272		
QY	248	GVVDNNGNGARVQRVVGSAAPASLIGISTGDIVTAVDGPAINSATAMADALNGHHPGVDIS 307		
Db	273	GVVDNNGNGARVQRVVGSAAPASLIGISTGDIVTAVDGPAINSATAMADALNGHHPGVDIS 332		
QY	308	VTWTKSGGTRTGNVTLAEGPPA 330		
Db	333	VTWTKSGGTRTGNVTLAEGPPA 355		

Db 436 APGATVALTFQDSPGSGSRVQVTLGK 461

RESULT 5
H86930
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002
C:Accession: H86930
R.; Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: H86930
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-382 <STO>
A:Cross-references: GB:AL450380; NID:gl3092536; PIDN:CAC29684.1; GSPDB:GNO0147
C:Genetics:
A:Gene: Mtu0176
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Query Match 26.3%; Score 447; DB 2; Length 382;
Best Local Similarity 37.2%; Pred. No. 1.8e-19;
Matches 125; Conservative 53; Mismatches 116; Indels 42; Gaps 12;

QY 18 ADFPALDPSAMVAQV-----PQVNVNTKLYNNVAGAGTGIVIDPNGVLTNNHVA 73
Db 59 ASVPAANM-PSGSVEQVAVKVPVVMLETLGRQSE--EGSGVILSADGLILTNHVA 115
QY 74 GA-----TDINAFSVGGQTYGVVDVVDYDTQDVAVLQLRGAGGLPSAAIG 119
Db 116 VAAKPGGGGGLSPKTVTFP---DGRASFTVVGADPTSDIAVVRVQSISGLTPTMG 172
QY 120 GG--VAVGEPVVMNGSGGQGTTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAA 177
Db 173 SSADLRVGPVAVVAGSPLAGI--VTSGIVSALNRPVSTTGE--SGNQNTVLDLQTDAA 229
QY 178 IQPGDAGGPVNVGLGVGMNTAA-----SDNFQLSQGGGFAIPIGQMAIAGQIRSG 231
Db 230 INFGSGGALVNMGGVLGVNSAATLGDAGSDGSGSIGLGFALPVDQAKRIADELIST 289
QY 232 GGSPTVHIGPTAFLGLGVVDNNGN--GARVQRVVGSAAPASLGISTGDIVTAVDGA PINSA 290
Db 290 G--KATH-----ASLGQVATDKGTPGAKVMDVVAGGAANAAPKGVVLTKVDDRLISSA 343
QY 291 TAMADALNGHHPGDVISVTWQTKSGGTRTGNVTLAE 326
Db 344 DALVAARVSKAPGDKVSLTYQDQSGSRRTVQVTLGK 379

RESULT 6
T45448
C:Species: Mycobacterium leprae
C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Dec-2002
C:Accession: T45448
R.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998
A:Reference number: D22967
A:Accession: T45448
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-452 <JAM>
A:Cross-references: EMBL:AL035500; PIDN:CAB36690.1
A:Experimental source: cosmid L373
C:Genetics:
A:Note: MLCB373.28
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Db 436 APGATVALTFQDSPGSGSRVQVTLGK 461

RESULT 5
H86930
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002
C:Accession: H86930
R.; Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: H86930
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-382 <STO>
A:Cross-references: GB:AL450380; NID:gl3092536; PIDN:CAC29684.1; GSPDB:GNO0147
C:Genetics:
A:Gene: Mtu0176
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Query Match 26.3%; Score 447; DB 2; Length 382;
Best Local Similarity 37.2%; Pred. No. 1.8e-19;
Matches 125; Conservative 53; Mismatches 116; Indels 42; Gaps 12;

QY 18 ADFPALDPSAMVAQV-----PQVNVNTKLYNNVAGAGTGIVIDPNGVLTNNHVA 73
Db 59 ASVPAANM-PSGSVEQVAVKVPVVMLETLGRQSE--EGSGVILSADGLILTNHVA 115
QY 74 GA-----TDINAFSVGGQTYGVVDVVDYDTQDVAVLQLRGAGGLPSAAIG 119
Db 116 VAAKPGGGGGLSPKTVTFP---DGRASFTVVGADPTSDIAVVRVQSISGLTPTMG 172
QY 120 GG--VAVGEPVVMNGSGGQGTTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAA 177
Db 173 SSADLRVGPVAVVAGSPLAGI--VTSGIVSALNRPVSTTGE--SGNQNTVLDLQTDAA 229
QY 178 IQPGDAGGPVNVGLGVGMNTAA-----SDNFQLSQGGGFAIPIGQMAIAGQIRSG 231
Db 230 INFGSGGALVNMGGVLGVNSAATLGDAGSDGSGSIGLGFALPVDQAKRIADELIST 289
QY 232 GGSPTVHIGPTAFLGLGVVDNNGN--GARVQRVVGSAAPASLGISTGDIVTAVDGA PINSA 290
Db 290 G--KATH-----ASLGQVATDKGTPGAKVMDVVAGGAANAAPKGVVLTKVDDRLISSA 343
QY 291 TAMADALNGHHPGDVISVTWQTKSGGTRTGNVTLAE 326
Db 344 DALVAARVSKAPGDKVSLTYQDQSGSRRTVQVTLGK 379

RESULT 6
T45448
C:Species: Mycobacterium leprae
C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Dec-2002
C:Accession: T45448
R.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998
A:Reference number: D22967
A:Accession: T45448
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-452 <JAM>
A:Cross-references: EMBL:AL035500; PIDN:CAB36690.1
A:Experimental source: cosmid L373
C:Genetics:
A:Note: MLCB373.28
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

F.182,224,305/Active site: His, Asp, Ser #status predicted

Query Match 26.3%; Score 447; DB 2; Length 452;
Best Local Similarity 37.2%; Pred. No. 2.2e-19;
Matches 125; Conservative 53; Mismatches 116; Indels 42; Gaps 12;

QY 18 ADFPALDPSAMVAQV-----PQVNVNTKLYNNVAGAGTGIVIDPNGVLTNNHVA 73
Db 129 ASVPAANM-PSGSVEQVAVKVPVVMLETLGRQSE--EGSGVILSADGLILTNHVA 185
QY 74 GA-----TDINAFSVGGQTYGVVDVVDYDTQDVAVLQLRGAGGLPSAAIG 119
Db 186 VAAKPGGGGGLSPKTVTFP---DGRASFTVVGADPTSDIAVVRVQSISGLTPTMG 242
QY 120 GG--VAVGEPVVMNGSGGQGTTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAA 177
Db 243 SSADLRVGPVAVVAGSPLAGI--VTSGIVSALNRPVSTTGE--SGNQNTVLDLQTDAA 299
QY 178 IQPGDAGGPVNVGLGVGMNTAA-----SDNFQLSQGGGFAIPIGQMAIAGQIRSG 231
Db 300 INFGSGGALVNMGGVLGVNSAATLGDAGSDGSGSIGLGFALPVDQAKRIADELIST 359
QY 232 GGSPTVHIGPTAFLGLGVVDNNGN--GARVQRVVGSAAPASLGISTGDIVTAVDGA PINSA 290
Db 360 G--KATH-----ASLGQVATDKGTPGAKVMDVVAGGAANAAPKGVVLTKVDDRLISSA 413
QY 291 TAMADALNGHHPGDVISVTWQTKSGGTRTGNVTLAE 326
Db 414 DALVAARVSKAPGDKVSLTYQDQSGSRRTVQVTLGK 449

RESULT 7
S74643
proteinase rhoA (EC 3.4.-.-) - Synecocystis sp. (strain PCC 6803)
N:Alternate names: protein all1679
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Dec-2002
R.; Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimoto, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
S.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S74643
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-394 <KAN>
A:Cross-references: EMBL:D90900; GB:AB001339; NID:gl651769; PIDN:BAAL16795.1; PID:gl651861
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: rhoA
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; trypr
C:Keywords: hydrolase; proteinase

Query Match 21.8%; Score 371; DB 2; Length 394;
Best Local Similarity 31.3%; Pred. No. 5.7e-15;
Matches 117; Conservative 50; Mismatches 111; Indels 96; Gaps 15;

QY 9 PPA-----LSQDRFADFALPLDPSAMVAQVQVNVNTK----- 44
Db 39 PPAPVITAAQSAPLTSFV-----AAVSRSGFAVVRIDTETVTRTRDPIILDDP 89
QY 45 -----LGYNNAV-----GAGTGIVIDPNGVLTNNHVIAGATDINAFSVGSGQTYGV 91
Db 90 FFQEFFGRSPVPRERRIAGQSGFIIDNSGIILTAHVVDGASKV-VVTLRDRGTFDG 148
QY 92 DVGYDRTQDVAVLQLRGAG-GLPSAAIG--GGVAVGEPVVMNGSGGQGTTPRAVGRV 148
Db 149 QVRGTDEVTDLAVVKIEPQGSALFPAFLGTSNLSQVGDWAIVAGNPGVLONT----- 200
QY 149 VALG--QTVQASDSLTGAETLNGLIQFDAAIQDGDAGGPVNVGLGVGMNTAASDNFQ 206

Db 201 VTIGIISTLGRSAQAQIPKRVFEFTQDAAINPNSGGPFLNAREVIGINTA---IR 256
QY 207 LSGGGGFAIPQIAQAIQIRSGGSPVTHGPTAFLGLGV---VNNGN----- 255
Db 257 ADATGIGFAIPQAKAIQNTLAAGTVEHPYIG-VQMNITVDQAQQNRNPNPFFIIP 315
QY 256 ---GARVQVVGSAASLGISTGDIYAVDGAIPINSATAM-----ADALN 298
Db 316 EVGILVMRVLPTPAERAGIRRGDVIVAVDGPISDGAIRQIVEQAGLNKALKLDLIR 375
QY 299 GHPGPDVISVTWQT 312
Db 376 GDRR---LSLTVQT 386

RESULT 8
AG2150
serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG2150
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG2150
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-407 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA74457.1; PID:ga7131851; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2758
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp

Query Match 21.5%; Score 365; DB 2; Length 407;
Best Local Similarity 31.2%; Pred. No. 1.3e-14;
Matches 109; Conservative 52; Mismatches 118; Indels 70; Gaps 11;
QY 28 SAMVAQVGPQVNTKLGYNNAV-----GAGTGIVD 60
Db 74 TAANRVGSAVRIDTTRTVDFLEDPFRFRFEGFQGLPPEQMRGLGSGFIID 133
QY 61 PNGVLTNNHVIAGATDINAFSGGQTYGVYDRTQDVAVLQRLGAGGLPSAAG- 119
Db 134 KSGILITNAHVVDKADRTV-RLKDGSRFDKQGVIDEVTDLAVVKINAGNSLFVAPLGS 192
QY 120 -GVAVGEPVVMNNSGGGGTTPRAGRVVAGL--QTVOASDLSLTGAETLNGLIQFPA 176
Db 193 SNNVQVGDMAIAGVNPGLGPDNT-----VTILGIVSTLKRSAQVIGITDKLDFIQTA 244
QY 177 AIOPGDAGGPVWNLGQVGMNTAAASDNFOLSGGQGFAPIQGAQIAQIRSGGSP 236
Db 245 AINPNSGGFLNDKKEVIGINTA---IFADAMGIGFALPIDKAKAIATQLERDQ--KV 298
QY 237 VHIGPTAFLGLGV-----VDNNG-----NGARVORVVGSAAPASLSTGSD 277
Db 299 AH-----PYLGVMQATLTPELAAQNNIDPNSAFAPAEVNGVVLVIRVVPNSPAAAGIRGD 354
QY 278 VITAVDGAIPNSATAMADALNGHHPGDVLSVTWTKSGGTRTGNVTLAE 326
Db 355 VILQVDCAITTAELQNVVNSRELQALQVRLQ-RGNQTQQLSVRTAE 402

RESULT 9
htrA-like serine protease (with PDZ domain) [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Dec-2002
C:Accession: H97199

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: H97199
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-433 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK80387.1; PID:gl5025449; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2433
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp

Query Match 20.6%; Score 350; DB 2; Length 433;
Best Local Similarity 31.6%; Pred. No. 1.1e-13;
Matches 100; Conservative 51; Mismatches 111; Indels 54; Gaps 10;
QY 25 LDPSAMVAQVGPQVNTKLGYNNAVAGTGTVIDPNGVLT 67
Db 122 LTVSQIVKVKVSPAVVGVSTKTITVTQNDFFSGSSNGSGSTQEGMSGIIFNNDGYILT 181
QY 68 NNHVIAGATDINAFSGGQTYGVYDRTQDVAVLQRLGAGGLPSAAGGVA--V 124
Db 182 NYHVIKADKT-AVILNNKKEVSAKVNYDEANDIAVIRKGTGFTVPGVAELSSASLNV 240
QY 125 GEPVVMNNSGGGGTTPRAGRVVAGLQTVQASDLSLTGAETLNGLIQFPAIQPDAG 184
Db 241 GDSVVAIGNPLGKEFLGTVTGTVVSAVNRVAVSE---GQKQT---YIQTDAINPNSG 294
QY 185 GPVTVNGLGQVGMNTAAASDNFOLSGGQ---GFAIPQGAQIAQIRSGGSPVTHIGP 241
Db 295 GPLVNSFGQVVGINS-----KISENGVEGIGFSIPIDTVKSKIONUSK-----P 339
QY 242 TAFILGLG--VVD-----NNGNGARVQVVGSAAPASLGTSTGDIYAVDGAIPINSATA 292
Db 340 ILMGLISGEAVDKSTASQHNIPQGVYIEQIQDFSSAQKAGMQVGDVITKEDGKKVTSTSD 399
QY 293 MADALNGHHPGDVLSV 308
Db 400 IDSIRKSHNSGDTVQV 415

RESULT 10
S77538
serine proteinase (EC 3.4.21.-) htrA - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein slr1204
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Dec-2002
C:Accession: S77538
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 103-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
S.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S77538
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-452 <KAN>
A:Cross-references: EMBL:D90905; GB:AB001339; NID:gl523260; PIDN:BAAL7385.1; PID:gl5246;
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: htrA
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp
C:Keywords: hydrolase; serine proteinase

Query Match 20.6%; Score 349; DB 2; Length 452;
Best Local Similarity 29.9%; Pred. No. 1.3e-13;
Matches 109; Conservative 60; Mismatches 113; Indels 82; Gaps 15;

QY 24 PLDPAMVAQV-----GPOVNTNTKLGYNNAV-----G 52
Db 112 PREPSNFVVDVVESTGPAVRINAQKTVKSPQPAENDFLQRFPGSQMPMPNERVORG 171
QY 53 AGTVIDPNDVVLNNHHVAGATDINAFSGVSGQTVGVVVDYDRTQDVAVLQURGAGG 112
Db 172 TGSQFIVSNDGKIFFTNAHVVDGADFTV-TLXDRGSRFGRVMGSDPSTDAVVKIE-AGD 229
QY 113 LPSAAIGGG--VAVEGPVVMGNSGGQGTTPRAVGRVVALGQTVQASDLSLTGABETLNG 170
Db 230 LFTVALGSDHQLQGEWAIRAINPGLDNT--VTGILSATGR--RSAD--IGVPDKRVE 283
QY 171 LIQFDAAIOPDAGGVVNGLGQVVMGNTAASDNFQLSGGQGFAPITGOAMAIAGQIRS 230
Db 284 FIOTDAAINPNSGGFLLNADQGVIMGTAIQNAQ-----GIGFAIPINKAQEIQAQLIA 339
QY 231 GGSSTVHHIGPTAFGLGLGVDDNN-----NGGARVQRVVSAPASLGI 273
Db 340 TG--KVHE-----AYLGIQVMTVTPBLSQIRQETGNIPVDKGVVIMQVMPNSPAAIAKL 393
QY 274 STGDVITAVDGAIPINSATAMADALNGHHPGDVISVTWQTKSGGTRTG---NVTLAGG--- 327
Db 394 EGGDVLQSLQGGQPVENAEQVSLVGLKLVGDEVEL-----GILNGQQQNLTVTIGALP 447
QY 328 --PP 329
Db 448 SAPP 451

RESULT 11

A:Accession: A82581
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-514 <SIM>
A:Cross-references: GB:AE004037; GB:AE003849; NID:99107394; PIDN:AAF95040.1; GSPDB:GN001
A:Experimental source: strain 945c
R:Simposon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.F.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Ishnako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: Xf2241
C:Superfamily: Helicobacter serine proteinase

Query Match 19.7%; Score 335; DB 2; Length 514;
Best Local Similarity 32.3%; Pred. No. 1e-12;
Matches 108; Conservative 49; Mismatches 107; Indels 70; Gaps 13;

QY 23 LPLDPAMVAQVGPQVNTNT-----KLGYNNAV----- 51
Db 54 LP-DFTQLVDVQVGPVNVNLTETWTKVKRGRIPLDNDIPBFRFRFGDPDQMPNQPRG 112
QY 52 -----GAGTGIVIDPNDVLTNNHHVAGATDINAFSGVSGQTVGVVDYDRTQ 100

Db 113 GQDDEGGIAGKMGSGFLISKDGVILTNHHVITGASEV-TIKLTDREFFAKIIGSDEQY 171
QY 101 DVAVLQURGAGGLPSAAIGGVAV--CEPVVANGNSGGQGTTPRAVGRVVALGQTVQAS 158
Db 172 DVALLKI-DANKLFTVRIIGDSSSLKSGQVVALGSPGLDHSVTA--GIVSALGRS----- 224
QY 159 DLSLTGABETLNGLTQFDAAIQPDAGGVVNGLGQVVMGNTAASDNFQLSQG--GQGFAL 216
Db 225 ---TSDQRYVPFIQTDVPIINQNSGGFLLNTRGEVIGIN---SQIFSAGGYMGISFAL 278
QY 217 PIGQAMAIAGQIRSGG-----GSPVHHIGPTAFGL---GLGVVDNNGGARVQRVVSAPAA 269
Db 279 PINLAINAABQIRKTKGVQSRMLGVEIGPIDALKAQGLGLPDS--RGALVNNIPPHSPAA 336
QY 270 SLGISTGDVITAVDGAIPINSATAMADALNGHHPG 303
Db 337 XAGIEVGDVIRSVNGKVISSFSDLPLPLIGMMPPG 370

RESULT 12

T35287
Probable secreted proteinase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Dec-2002
C:Accession: T35287
R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21574
A:Accession: T35287
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-362 <SSE>
A:Cross-references: EMBL:AL096872; PIDN:CAB51255.1; GSPDB:GN00070; SCOEDB:SC5F7.30
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC5F7.30
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp

Query Match 19.6%; Score 332.5; DB 2; Length 362;
Best Local Similarity 33.6%; Pred. No. 9.6e-13;
Matches 112; Conservative 46; Mismatches 138; Indels 37; Gaps 12;

QY 8 APPALSDQRADFPALPLDPSAMVAQVGPQVNTNTKLGYNNAVAGTGIVIDPNGVLT 67
Db 52 APPRASE-----LEADYERVIKVLPSVVQIA-----GDSLSGGVYDDKGHVVT 97
QY 68 NNHVIAGATDINAFSGVSGQTVG---VDVVGVDYDRTQDVAVLQL-RGAGGLPSA--AIGGG 121
Db 98 NAHVVG---DAKSPRVTTARTEGALTAKLVSSYPEQDLAVIKLDKVPKGMRAARFADSAK 154
QY 122 VAVEGPVVMGNSGGQGTTPRAVGRVVALGQTVQASDLSLTGABETLNGLIQPDAAIQPG 181
Db 155 VEVGQIVLVMGSPILGLSS--VTQGIYSATGRVTTEGSGGGGTGATIANMVQTSAAINPG 212
QY 182 DAGGEPVNGLGQVVMGNTAASDNFQLSQG--GQGFAPITGOAMAIAGQIRSGG---GSP 235
Db 213 NSGALVNLGDQVIGITFLAATDPLGDSAPAGLGAIPASMTVTAVAGIVRDGKVTDSG 272
QY 236 TVHHIGPTAFGLGVVDNNGN--GARVQRVVSAPASLGI-STGDTVITAVDGAIPINSATAM 293
Db 273 RAALGITA---RTVVDSDSYRPAAGAAVVEVSDGGAADAGLRPGDVLVKLGDITDITITSL 329
QY 294 ADALNGHHPGDVISVTWQTKSGGTRTGNVTLAE 326
Db 330 SEALASWRPGDRTKVTY-TRDGKHTAEVTLGE 361

RESULT 13

AB2057
serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C;Accession: AB2057
 R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena* sp. strain PCC 7120
 A;Reference number: AB1807; MUID:21595285; PMID:11759840
 A;Accession: AB2057
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-416 <KUR>
 A;Cross-references: GB:BA000019; PIDN:BA073707.1; PID:gl7131098; GSPDB:GN00179
 A;Experimental source: strain PCC 7120
 C;Genetics:
 C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try
 Query Match 19.4%; Score 330; DB 2; Length 416;
 Best Local Similarity 28.7%; Pred. No. 1.6e-12;
 Matches 100; Conservative 64; Mismatches 126; Indels 58; Gaps 12;
 26 DPS---AMVAQGPQVNVNINIKLYNNV-----GAGT 55
 77 DNFVVGQVGVAVRIDSARTVTSRVNENFDFRFGDGVPAQPRQVRGSGS 136
 56 GIVDPNGVLTNNHVIAGATDINAFSGSGTYGVVGVYDRTQDVAVLQRLGAGLPS 115
 137 GFTISSGQLTNAHVVGDGADVTV-TLKDGRFDFGKVLGEDPVTVDVAVIQI-NANNLPT 194
 116 AAIQGGGVAV--GPFVAMGSGGGTTPRAVGRVVALGQTVQASDSLTCGAETLNLGLI 173
 195 VAVGNSVLPQGEAIVAGNPLGNNS--VTSGIISATGR---SGSDIGASDKRVDYIQ 248
 174 FDAIQDQDAGGVPVNLGVGVGVMYTAASDNFQSGGQGFAPICQAMAIAGQIRSGG 233
 249 TDAIINPGSGGELLNARGQVIGMGTATIQGAQ---GLGFAIPINTVQKVSQELITQK 304
 234 SPVTHIG-PTAFILGLGVD--NNGGARVQ-----RVVGSAPAASLGTSTGDTVIT 280
 305 VDPHVLGVQWATITPQVKERINERFGRINITADRGVLLVRIVPGSPAANAGLRFGDI 364
 281 AVDGAPINSATAMADALNGHHHPGDVTSVTWTKSGGTRTGNVTLAGRP 328
 365 SINQSVTTVEQVKIVENSIGQPLQIQIE-RNGQTTQVNVSPAPLP 411
 RESULT 14
 B71284
 probable periplasmic serine proteinase DO (htrA-1) - syphilis spirochete
 C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
 C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 09-Dec-2002
 C;Accession: B71284
 R;Praser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
 rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD
 chey, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
 Science 281, 375-388, 1998
 A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
 A;Reference number: A71250; MUID:98332770; PMID:9665876
 A;Accession: B71284
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-398 <COL>
 A;Cross-references: GB:AE001248; GB:AE000520; NID:93323074; PIDN:AAC65740.1; PID:9332308
 A;Experimental source: strain Nichols
 C;Genetics:
 A;Gene: TP0773
 C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try
 Query Match 19.4%; Score 329; DB 2; Length 398;
 Best Local Similarity 32.1%; Pred. No. 1.7e-12;
 Matches 103; Conservative 43; Mismatches 121; Indels 54; Gaps 11;
 29 AMVAQGPQVNVNINIKL-GYNNV-----GAGTGIVDPNGVLTNNHVIAGATDINA 80

Db 72 AVYSANEAVNIITTEVMGVNMFLEPVLPGSGSGAIIDARGVYLTNTHTVIEGASKI-Y 130
 QY 81 FSVSGGQTYGVVGVYDRTQDVAVLQRLGAGLPSAIIQGG-----VAVGPPVAMGNSGG 136
 Db 131 LSLHDGSQYKATVGVVDRENDLAVLVKVPSPGAEITVIRGSSRNLDVGQKVLAINPFG 190
 QY 137 QGGTPRAVGRVVALGQTVQASDSLTCGAETLNLGLIOFDAIQDQDAGGPPVNLGQVVG 196
 Db 191 LART--ITVGVSALARPIONKSGI-----IRNMIQTDAINPNSGGPILDTQGRMIG 242
 QY 197 MNTAASDNFQSGGQGFAPICQAMAI-----GOIRSG-----GGSPTVHIGTAFGL 247
 Db 243 INTVIYST-SGSSGVGFAPVDTAKRIVSELIRYGVRRGKIDAEILVQNVASTAHVAQL 301
 QY 248 GVDNNGNGARVORVVGSAASLGIS-----TGDIVTAVDGA PIN 288
 Db 302 TV---GKLLVSQVKGSPAAQAGLRGTTAVRYGLRRAAVIYLGDIVTIDNQPA 357
 QY 289 SATAMADALNGHHHPGDVTSVT 309
 Db 358 NLSDYSVLEDKKXPDDEVRT 378
 RESULT 15
 E75357
 probable periplasmic serine proteinase Do - Deinococcus radiodurans (strain R1)
 C;Species: Deinococcus radiodurans
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Dec-2002
 C;Accession: E75357
 R;White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; F
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A;Reference number: A75250; MUID:20036896; PMID:10567266
 A;Accession: E75357
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-441 <WHI>
 A;Cross-references: GB:AE002017; GB:AE000513; NID:g6459527; PIDN:AAF11312.1; PID:g645953
 A;Experimental source: strain R1
 C;Genetics:
 A;Gene: DR1756
 A;Map position: 1
 C;Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try
 Query Match 19.4%; Score 329; DB 2; Length 441;
 Best Local Similarity 29.3%; Pred. No. 1.9e-12;
 Matches 106; Conservative 50; Mismatches 110; Indels 96; Gaps 11;
 QY 14 QDREAFDPPALPLDPSAMVAQVGPQVNVNINIKLYNNVAVGAGTGIVDPNGVLTNNHVI 73
 Db 115 QDQGF--ALPEDOG-----GAPATGSGGFFVNAQGDIVTNNHVD 155
 QY 74 GATINAFSGSGGTYGVVGVYDRTQDVAVLQRLGAGLPSAII-----GGVAVGE 126
 Db 156 GASDITIRLHGKQTKYAKVIARAPDYDLAIR---AEGVPRASIKPLPLGDSDKLDVGL 212
 QY 127 PVVAMG-----NSGGGGGTTPRAVGRVVALGQTVQASDSLTCGAETLNLGLIOFDA 177
 Db 213 KALANGAPFNLDPSVSGIISLIERQVP-----VGSREVSPQVIQTDA 256
 QY 178 IQPDAGGPPVNLGQVGVGMNTAASDNFQSGGQGFAPICQAMAIAGQIRSGGSPV 237
 Db 257 INPNSGGLPLSSAGQVIGVNTQILTGGAGQSAGVGAIPINTVXRLLPLOAQKG--- 312
 QY 238 HIGTAFILGLGVDNN-----NGCARVORVVGSAASLG----- 272
 Db 313 --GVSPSPSGVVFSDUSSIPPOOLKAGLPPSSGALLQKTYPGSPAAAGLRGNNKLS 370
 QY 273 -----IST-GDIVTAVDGA PIN SATAMADALNGHHHPGDVTSVTWTKSGGTRTGNV 322

Db 371 LPSAQGTSSISTDGLITAVNGQPLEDAGSLOEAVLATGEGQPLRLTVR-RGGKTREVEV 429

QY 323 TL 324

Db 430 TL 431

Search completed: June 22, 2004, 17:24:49
Job time : 10.233 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 30.5644 Seconds
(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-6
Perfect score: 1698
Sequence: 1 MHHEHHAPPALSDQRFADF.....QTKSGGTRGNVTLAEGPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/FCT_NEW PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/FCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1698	100.0	330	12	US-09-886-349A-6
2	1698	100.0	330	14	US-10-098-732A-6
3	1698	100.0	330	15	US-10-369-983-20
4	1698	100.0	723	15	US-10-369-983-2
5	1698	100.0	1010	15	US-10-369-983-4
6	1695	99.8	330	12	US-09-886-349A-4
7	1695	99.8	330	14	US-10-098-732A-4
8	1695	99.8	330	15	US-10-369-983-19
9	1642	96.7	355	9	US-09-712-363-161
10	1631	96.1	355	12	US-09-886-349A-2
11	1631	96.1	355	14	US-10-193-002-80
12	1631	96.1	355	14	US-10-084-843-79
13	1631	96.1	355	14	US-10-098-732A-2
14	1000.5	58.9	1016	15	US-10-369-983-18
15	992	58.4	1154	15	US-10-369-983-16

16	991.5	58.4	825	15	US-10-369-983-14	Sequence 14, Appl
17	990.5	58.3	813	15	US-10-369-983-15	Sequence 15, Appl
18	989	58.2	1022	15	US-10-369-983-17	Sequence 17, Appl
19	987	58.1	729	12	US-09-886-349A-18	Sequence 18, Appl
20	987	58.1	729	14	US-10-098-732A-18	Sequence 18, Appl
21	987	58.1	729	15	US-10-369-983-22	Sequence 22, Appl
22	984.5	58.0	875	15	US-10-369-983-13	Sequence 13, Appl
23	984	58.0	596	9	US-09-287-849-26	Sequence 26, Appl
24	984	58.0	596	12	US-09-886-349A-20	Sequence 20, Appl
25	984	58.0	596	14	US-10-359-460-26	Sequence 26, Appl
26	984	58.0	596	14	US-10-098-732A-20	Sequence 20, Appl
27	984	58.0	729	9	US-09-287-849-2	Sequence 2, Appl
28	984	58.0	729	12	US-09-886-349A-16	Sequence 16, Appl
29	984	58.0	729	14	US-10-359-460-2	Sequence 2, Appl
30	984	58.0	729	14	US-10-098-732A-16	Sequence 16, Appl
31	984	58.0	729	15	US-10-369-983-21	Sequence 21, Appl
32	984	58.0	729	15	US-10-359-459-2	Sequence 2, Appl
33	984	58.0	930	14	US-10-098-732A-65	Sequence 12, Appl
34	984	58.0	930	15	US-10-369-983-12	Sequence 12, Appl
35	978	57.6	195	12	US-09-886-349A-8	Sequence 8, Appl
36	978	57.6	195	14	US-10-098-732A-8	Sequence 8, Appl
37	676	39.8	231	9	US-09-287-849-28	Sequence 28, Appl
38	676	39.8	231	14	US-10-359-460-28	Sequence 28, Appl
39	670	39.5	132	9	US-09-759-143-819	Sequence 819, App
40	670	39.5	132	9	US-09-759-143-848	Sequence 848, App
41	670	39.5	132	9	US-09-780-669-819	Sequence 819, App
42	670	39.5	132	9	US-09-780-669-848	Sequence 848, App
43	670	39.5	132	9	US-09-822-827-819	Sequence 819, App
44	670	39.5	132	9	US-09-822-827-848	Sequence 848, App
45	670	39.5	132	9	US-09-895-793-819	Sequence 819, App

ALIGNMENTS

RESULT 1
US-09-886-349A-6
; Sequence 6, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; TYPE: PRT
; LENGTH: 330
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Ra35FLMutSA
US-09-886-349A-6

Query Match	100.0%;	Score 1698;	DB 12;	Length 330;
Best Local Similarity	100.0%;	Pred. No. 1.5e-121;	Mismatches 0;	Indels 0;
Matches 330;	Conservative 0;			
QY	1	MHHEHHAPPALSDQRFADFALP	LDPSAMVAQVGPQVNVINTKLGYNNAVGAGTGIVID	60
DB	1	MHHEHHAPPALSDQRFADFALP	LDPSAMVAQVGPQVNVINTKLGYNNAVGAGTGIVID	60
QY	61	PNGVVLNNHVIAGATDINAFSVGSQGYGVDDVGYGRTQDVAVLQLRGAGLPSAIGG		120
DB	61	PNGVVLNNHVIAGATDINAFSVGSQGYGVDDVGYGRTQDVAVLQLRGAGLPSAIGG		120

QY 121 GVAVGEPVVMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
Db 121 GVAVGEPVVMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
QY 181 GDAGGPVVGNGLGQVVGWNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
Db 181 GDAGGPVVGNGLGQVVGWNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVVGSAPASLSIGTSGDVTAVDGPINSATAMADALNGH 300
Db 241 PTAFLGLGVVDNNGNGARVQVVVGSAPASLSIGTSGDVTAVDGPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 2

US-10-098-732A-6
; Sequence 6, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 6
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Ra35FLMutSA
US-10-098-732A-6

Query Match 100.0%; Score 1698; DB 14; Length 330;
Best Local Similarity 100.0%; Pred. No. 1.5e-121;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNVAVGAGTGIVID 60
Db 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNVAVGAGTGIVID 60
QY 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120
Db 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120
QY 121 GVAVGEPVVMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
Db 121 GVAVGEPVVMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
QY 181 GDAGGPVVGNGLGQVVGWNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
Db 181 GDAGGPVVGNGLGQVVGWNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVVGSAPASLSIGTSGDVTAVDGPINSATAMADALNGH 300
Db 241 PTAFLGLGVVDNNGNGARVQVVVGSAPASLSIGTSGDVTAVDGPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 3

US-10-369-983-20
; Sequence 20, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 20
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: mutated
; OTHER INFORMATION: MTB32AMutSA (Ra35 mutSA)
US-10-369-983-20

Query Match 100.0%; Score 1698; DB 15; Length 330;
Best Local Similarity 100.0%; Pred. No. 1.5e-121;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNVAVGAGTGIVID 60
Db 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNVAVGAGTGIVID 60
QY 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120
Db 61 PNGVVLNNHVIAGATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGG 120
QY 121 GVAVGEPVVMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
Db 121 GVAVGEPVVMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
QY 181 GDAGGPVVGNGLGQVVGWNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
Db 181 GDAGGPVVGNGLGQVVGWNTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVVGSAPASLSIGTSGDVTAVDGPINSATAMADALNGH 300
Db 241 PTAFLGLGVVDNNGNGARVQVVVGSAPASLSIGTSGDVTAVDGPINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 4

US-10-369-983-2
; Sequence 2, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2

```
; LENGTH: 723
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: mutated
; OTHER INFORMATION: MTB32-MTB39f fusion protein (MTB32MutSA)
US-10-369-983-2

Query Match      100.0%; Score 1698; DB 15; Length 723;
Best Local Similarity 100.0%; Pred. No. 3.9e-121;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQGVQVNVNNTKLGYNNAVAGTGIVID 60
Db 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQGVQVNVNNTKLGYNNAVAGTGIVID 60
QY 61 PNGVLTNNHVIAGATDINAFSGSGQTYGVDVVGVDRTQDVAVLQIRGAGLPSAAIGG 120
Db 61 PNGVLTNNHVIAGATDINAFSGSGQTYGVDVVGVDRTQDVAVLQIRGAGLPSAAIGG 120
QY 121 GVAAGPVPVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQP 180
Db 121 GVAAGPVPVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQP 180
QY 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
Db 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSLGISTGDTVITAVDGA PINSATAMADALNGH 300
Db 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSLGISTGDTVITAVDGA PINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 5
US-10-369-983-4
; Sequence 4, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB-102F fusion
US-10-369-983-4

Query Match      100.0%; Score 1698; DB 15; Length 1010;
Best Local Similarity 100.0%; Pred. No. 5.9e-121;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQGVQVNVNNTKLGYNNAVAGTGIVID 60
Db 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQGVQVNVNNTKLGYNNAVAGTGIVID 60
QY 61 PNGVLTNNHVIAGATDINAFSGSGQTYGVDVVGVDRTQDVAVLQIRGAGLPSAAIGG 120
Db 61 PNGVLTNNHVIAGATDINAFSGSGQTYGVDVVGVDRTQDVAVLQIRGAGLPSAAIGG 120
```

```
Db 61 PNGVLTNNHVIAGATDINAFSGSGQTYGVDVVGVDRTQDVAVLQIRGAGLPSAAIGG 120
QY 121 GVAAGPVPVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQP 180
Db 121 GVAAGPVPVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQP 180
QY 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
Db 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSLGISTGDTVITAVDGA PINSATAMADALNGH 300
Db 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSLGISTGDTVITAVDGA PINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 6
US-09-886-349A-4
; Sequence 4, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB32A (Ra35 mature)
US-09-886-349A-4

Query Match      99.8%; Score 1695; DB 12; Length 330;
Best Local Similarity 99.7%; Pred. No. 2.6e-121;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQGVQVNVNNTKLGYNNAVAGTGIVID 60
Db 1 MHHHHHAPPALSDQRFADFPALPLDPSAMVAQGVQVNVNNTKLGYNNAVAGTGIVID 60
QY 61 PNGVLTNNHVIAGATDINAFSGSGQTYGVDVVGVDRTQDVAVLQIRGAGLPSAAIGG 120
Db 61 PNGVLTNNHVIAGATDINAFSGSGQTYGVDVVGVDRTQDVAVLQIRGAGLPSAAIGG 120
QY 121 GVAAGPVPVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQP 180
Db 121 GVAAGPVPVAMGNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQP 180
QY 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
Db 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSLGISTGDTVITAVDGA PINSATAMADALNGH 300
Db 241 PTAFLGLGVVDNNGNGARVQVVGSAAPASLSLGISTGDTVITAVDGA PINSATAMADALNGH 300
QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
Db 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
```

```
RESULT 7
US-10-098-732A-4
; Sequence 4, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Lelishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT FILING DATE: 2003-04-29
; PRIOR FILING DATE: 2003-04-29
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: Ra35 mature
US-10-098-732A-4
Query Match 99.8%; Score 1695; DB 14; Length 330;
Best Local Similarity 99.7%; Pred. No. 2.6e-121;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHAPPALSDRFPALPLDPSAMVAQVGPVNNINIKLGYNNAVGAGTGIVID 60
DB 1 MHHHHHAPPALSDRFPALPLDPSAMVAQVGPVNNINIKLGYNNAVGAGTGIVID 60
QY 61 PNGVLTNNHVIAGATDINAFSVSGGQTYGVVGVYDRTQDVAVLQIRGAGGLPSAAIGG 120
DB 61 PNGVLTNNHVIAGATDINAFSVSGGQTYGVVGVYDRTQDVAVLQIRGAGGLPSAAIGG 120
QY 121 GVAVGEPVWANGSGGGGTTPRVPGRVVALGQTVQASDSLTAETLNGLIQFDDAIIQ 180
DB 121 GVAVGEPVWANGSGGGGTTPRVPGRVVALGQTVQASDSLTAETLNGLIQFDDAIIQ 180
QY 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGQFAIPIGQMAIAGQIRSGGSPVTHIG 240
DB 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGQFAIPIGQMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVNDNNGARVQVWGSAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
DB 241 PTAFLGLGVNDNNGARVQVWGSAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
QY 301 HPGDVISVTWTKSGGTRTGNVTLAEGPPA 330
DB 301 HPGDVISVTWTKSGGTRTGNVTLAEGPPA 330
RESULT 8
US-10-369-983-19
; Sequence 19, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 161
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: Ra35 mature
US-10-098-732A-4
Query Match 99.8%; Score 1695; DB 14; Length 330;
Best Local Similarity 99.7%; Pred. No. 2.6e-121;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHAPPALSDRFPALPLDPSAMVAQVGPVNNINIKLGYNNAVGAGTGIVID 60
DB 1 MHHHHHAPPALSDRFPALPLDPSAMVAQVGPVNNINIKLGYNNAVGAGTGIVID 60
QY 61 PNGVLTNNHVIAGATDINAFSVSGGQTYGVVGVYDRTQDVAVLQIRGAGGLPSAAIGG 120
DB 61 PNGVLTNNHVIAGATDINAFSVSGGQTYGVVGVYDRTQDVAVLQIRGAGGLPSAAIGG 120
QY 121 GVAVGEPVWANGSGGGGTTPRVPGRVVALGQTVQASDSLTAETLNGLIQFDDAIIQ 180
DB 121 GVAVGEPVWANGSGGGGTTPRVPGRVVALGQTVQASDSLTAETLNGLIQFDDAIIQ 180
QY 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGQFAIPIGQMAIAGQIRSGGSPVTHIG 240
DB 181 GDAGGPVNVGLGVVGMNTAASDNFQLSQGGQFAIPIGQMAIAGQIRSGGSPVTHIG 240
QY 241 PTAFLGLGVNDNNGARVQVWGSAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
DB 241 PTAFLGLGVNDNNGARVQVWGSAPASLSIGTGDVITAVDGPINSATAMADALNGH 300
QY 301 HPGDVISVTWTKSGGTRTGNVTLAEGPPA 330
DB 301 HPGDVISVTWTKSGGTRTGNVTLAEGPPA 330
RESULT 9
US-09-712-363-161
; Sequence 161, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT FILING DATE: 2000-11-13
; PRIOR FILING DATE: 2000-11-13
; PRIOR FILING DATE: 2000-01-28
; PRIOR FILING DATE: 2000-01-28
; PRIOR FILING DATE: 2000-02-01
; PRIOR FILING DATE: 2000-02-01
; PRIOR FILING DATE: 1999-01-29
; PRIOR FILING DATE: 1999-01-29
; PRIOR FILING DATE: 1999-05-14
; PRIOR FILING DATE: 1999-05-14
; PRIOR FILING DATE: 1999-05-14
; PRIOR FILING DATE: 1999-11-12
; PRIOR FILING DATE: 1999-11-12
; PRIOR FILING DATE: 1999-11-12
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 161
```

```
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-161

Query Match
Best Local Similarity 96.7%; Score 1642; DB 9; Length 355;
Matches 322; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGAGTIVIDPVGWVLT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGAGTIVIDPVGWVLT 92
QY 68 NNHVIAGATDINAFSGSGQTGVGVVGYDRTQDVAVLQRLRGAGGLPSAAIGGVAVGEP 127
DB 93 NNHVIAGATDINAFSGSGQTGVGVVGYDRTQDVAVLQRLRGAGGLPSAAIGGVAVGEP 152
QY 128 VVAMNSGGGGTTPRAVGRVVALGQTVAQSDSLTGAEETLNGLIQFDDAALQPGDAGGPV 187
DB 153 VVAMNSGGGGTTPRAVGRVVALGQTVAQSDSLTGAEETLNGLIQFDDAALQPGDAGGPV 212
QY 188 VVAMNSGGGGTTPRAVGRVVALGQTVAQSDSLTGAEETLNGLIQFDDAALQPGDAGGPV 247
DB 213 VVAMNSGGGGTTPRAVGRVVALGQTVAQSDSLTGAEETLNGLIQFDDAALQPGDAGGPV 272
QY 248 GVVDDNNGGARVQVVGSAFASLGIPIGQMAIAGQIRSGGSPVHIGPTAFGL 307
DB 273 GVVDDNNGGARVQVVGSAFASLGIPIGQMAIAGQIRSGGSPVHIGPTAFGL 332
QY 308 VTWQKSGGTRTGNVTLAEGPPA 330
DB 333 VTWQKSGGTRTGNVTLAEGPPA 355

RESULT 11
US-10-193-002-80
; Sequence 80, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10193,002
; FILING DATE: 10-Jul-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 80:
US-10-193-002-80

Query Match
Best Local Similarity 96.1%; Score 1631; DB 14; Length 355;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-161

Query Match
Best Local Similarity 99.7%; Score 1642; DB 9; Length 355;
Matches 322; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGAGTIVIDPVGWVLT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGAGTIVIDPVGWVLT 92
QY 68 NNHVIAGATDINAFSGSGQTGVGVVGYDRTQDVAVLQRLRGAGGLPSAAIGGVAVGEP 127
DB 93 NNHVIAGATDINAFSGSGQTGVGVVGYDRTQDVAVLQRLRGAGGLPSAAIGGVAVGEP 152
QY 128 VVAMNSGGGGTTPRAVGRVVALGQTVAQSDSLTGAEETLNGLIQFDDAALQPGDAGGPV 187
DB 153 VVAMNSGGGGTTPRAVGRVVALGQTVAQSDSLTGAEETLNGLIQFDDAALQPGDAGGPV 212
QY 188 VVAMNSGGGGTTPRAVGRVVALGQTVAQSDSLTGAEETLNGLIQFDDAALQPGDAGGPV 247
DB 213 VVAMNSGGGGTTPRAVGRVVALGQTVAQSDSLTGAEETLNGLIQFDDAALQPGDAGGPV 272
QY 248 GVVDDNNGGARVQVVGSAFASLGIPIGQMAIAGQIRSGGSPVHIGPTAFGL 307
DB 273 GVVDDNNGGARVQVVGSAFASLGIPIGQMAIAGQIRSGGSPVHIGPTAFGL 332
QY 308 VTWQKSGGTRTGNVTLAEGPPA 330
DB 333 VTWQKSGGTRTGNVTLAEGPPA 355

RESULT 10
US-09-886-349A-2
; Sequence 2, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB32A (Ra35FL)
US-09-886-349A-2

Query Match
Best Local Similarity 99.1%; Score 1631; DB 12; Length 355;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGAGTIVIDPVGWVLT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGAGTIVIDPVGWVLT 92
QY 68 NNHVIAGATDINAFSGSGQTGVGVVGYDRTQDVAVLQRLRGAGGLPSAAIGGVAVGEP 127
DB 93 NNHVIAGATDINAFSGSGQTGVGVVGYDRTQDVAVLQRLRGAGGLPSAAIGGVAVGEP 152
```

QY 8 APPALSODRFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGVIVDPNGVLT 67
DB 33 APPALSODRFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGVIVDPNGVLT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGLPSAAIGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGLPSAAIGGVAVGEP 152
QY 128 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDAGGPV 187
DB 153 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDAGGPV 212
QY 188 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDAGGPV 247
DB 213 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDAGGPV 272
QY 248 GVVDNNGNGARVQVGVSGAPASLIGSTGDTAVDGPINSATAMADALNGHHHPGDVIS 307
DB 273 GVVDNNGNGARVQVGVSGAPASLIGSTGDTAVDGPINSATAMADALNGHHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 12

US-10-084-843-79
; Sequence 79, Application US/10084843
; Publication No. US20030143243A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
Skeiky, Yasir A.W.
Dillon, David C.
Campos-Neto, Antonio
Houghton, Raymond
Vedvick, Thomas S.
Twardzik, Daniel R.
Lodes, Michael J.
Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.411C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 79:

SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-10-084-843-79

Query Match 96.1%; Score 1631; DB 14; Length 355;
Best Local Similarity 99.1%; Pred. No. 2.1e-116;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSODRFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGVIVDPNGVLT 67
DB 33 APPALSODRFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGVIVDPNGVLT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGLPSAAIGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGLPSAAIGGVAVGEP 152
QY 128 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDAGGPV 187
DB 153 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDAGGPV 212
QY 188 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDAGGPV 247
DB 213 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDAGGPV 272
QY 248 GVVDNNGNGARVQVGVSGAPASLIGSTGDTAVDGPINSATAMADALNGHHHPGDVIS 307
DB 273 GVVDNNGNGARVQVGVSGAPASLIGSTGDTAVDGPINSATAMADALNGHHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 13

US-10-098-732A-2
; Sequence 2, Application US/10098732A
; Publication No. US20030175294A1

GENERAL INFORMATION:

APPLICANT: Skeiky, Yasir
APPLICANT: Brannon, Mark
APPLICANT: Guderman, Jeffrey
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
FILE REFERENCE: 014058-012010US
CURRENT APPLICATION NUMBER: US/10/098,732A
CURRENT FILING DATE: 2003-04-29
PRIOR FILING DATE: 2001-03-13
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 355
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: MTE32A (Ra35FL)

US-10-098-732A-2

Query Match 96.1%; Score 1631; DB 14; Length 355;
Best Local Similarity 99.1%; Pred. No. 2.1e-116;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSODRFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGVIVDPNGVLT 67
DB 33 APPALSODRFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGVIVDPNGVLT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGLPSAAIGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGLPSAAIGGVAVGEP 152
QY 128 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDAGGPV 187

Db 153 VVAMGNSGGGCTPRAPGRRVVALGQTVQASDSLGTGABETLNGLIQFDAAIQPGDSGGPV 212
Qy 188 VNGLGQVVGWNTAAADNFQLSGGGQGFALPIQAMAIAGQIRSGGSPVHIGPTAFGL 247
Db 213 VNGLGQVVGWNTAAADNFQLSGGGQGFALPIQAMAIAGQIRSGGSPVHIGPTAFGL 272
Qy 248 GVVDDNNGGARVQVVVGAPASLSGTSTGDTVITAVDGPINSATAMADALNGHHPGDVIS 307
Db 273 GVVDDNNGGARVQVVVGAPASLSGTSTGDTVITAVDGPINSATAMADALNGHHPGDVIS 332
Qy 308 VTWQKSGGTRTGNVTLAEGPPA 330
Db 333 VNWQKSGGTRTGNVTLAEGPPA 355

RESULT 14
US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85b)
US-10-369-983-18

Query Match 58.9%; Score 1000.5; DB 15; Length 1016;
Best Local Similarity 85.5%; Pred. No. 9.5e-68;
Matches 206; Conservative 7; Mismatches 13; Indels 15; Gaps 4;

Qy 8 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGIVIDPNGVLT 67
Db 535 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGIVIDPNGVLT 594
Qy 68 NNHVIAGATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGGLPSAAICGGVAVGEP 127
Db 595 NNHVIAGATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGGLPSAAICGGVAVGEP 654
Qy 128 VVAMGNSGGGCTPRAPGRRVVALGQTVQASDSLGTGABETLNGLIQFDAAIQPGDSGGPV 187
Db 655 VVAMGNSGGGCTPRAPGRRVVALGQTVQASDSLGTGABETLNGLIQFDAAIQPGDSGGPV 714
Qy 188 VNGLGQVVGWNTAAADNFQLSGGGQGFALPIQAMAIAGQIRSGGSPVHIGPTAFGL 247
Db 715 VNGLGQVVGWNTAAADNFQLSGGGQGFALPIQAMAIAGQIRSGGSPVHIGPTAFGL 296
Qy 239 I 239
Db 769 L 769

RESULT 15
US-10-369-983-16
; Sequence 16, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff

; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB114F (MTB72F-mTCC#2)
US-10-369-983-16
Query Match 58.4%; Score 992; DB 15; Length 1154;
Best Local Similarity 72.9%; Pred. No. 4.9e-67;
Matches 212; Conservative 14; Mismatches 37; Indels 28; Gaps 4;
Qy 8 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGIVIDPNGVLT 67
Db 535 APPALSQDRFADFPALPLDPSAMVAQVGVVNTKLGYNNAVAGTGIVIDPNGVLT 594
Qy 68 NNHVIAGATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGGLPSAAICGGVAVGEP 127
Db 595 NNHVIAGATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGGLPSAAICGGVAVGEP 654
Qy 128 VVAMGNSGGGCTPRAPGRRVVALGQTVQASDSLGTGABETLNGLIQFDAAIQPGDSGGPV 187
Db 655 VVAMGNSGGGCTPRAPGRRVVALGQTVQASDSLGTGABETLNGLIQFDAAIQPGDSGGPV 714
Qy 188 VNGLGQVVGWNTAAADNFQLSGGGQGFALPIQAMAIAGQIRSGGSPVHIGPTAFGL 247
Db 715 VNGLGQVVGWNTAAADNFQLSGGGQGFALPIQAMAIAGQIRSGGSPVHIGPTAFGL 296
Qy 248 GVVDDNNGGARVQVVVGAPASLSGTSTGDTVITAVDGPINSATAMADALNGHHPGDVIS 307
Db 766 GVVDDNNGGARVQVVVGAPASLSGTSTGDTVITAVDGPINSATAMADALNGHHPGDVIS 332

Search completed: June 22, 2004, 18:07:52
Job time : 30.5644 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:44:25 ; Search time 40.5137 Seconds
(without alignments)
2301.458 Million cell updates/sec

Title: US-09-886-349A-6
Perfect score: 1698
Sequence: 1 MHHHHHAPPALSQDRPADF.....QTKSGTGTGNVTLAEGPPA 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04: *
1: Geneseqp1980s: *
2: Geneseqp1990s: *
3: Geneseqp2000s: *
4: Geneseqp2001s: *
5: Geneseqp2002s: *
6: Geneseqp2003as: *
7: Geneseqp2003bs: *
8: Geneseqp2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1698	100.0	330	5 AAE29703	Aae29703 Mycobacte
2	1698	100.0	330	5 AAE17567	Aae17567 Mycobacte
3	1698	100.0	330	7 ADA26372	Ada26372 Mycobacte
4	1698	100.0	723	7 ADA26354	Ada26354 Mycobacte
5	1698	100.0	1010	7 ADA26356	Ada26356 Mycobacte
6	1695	99.8	330	5 AAE29702	Aae29702 Mycobacte
7	1695	99.8	330	5 AAE17566	Aae17566 Mycobacte
8	1688	99.4	330	7 ADA26371	Ada26371 Mycobacte
9	1642	96.7	355	2 AAY05000	Aay05000 Mycobacte
10	1642	96.7	355	4 AAO22137	Aao22137 Mycobacte
11	1642	96.7	355	4 AAG81110	Aag81110 Mycobacte
12	1642	96.7	379	2 AAY04830	Aay04830 Mycobacte
13	1631	96.1	355	2 AAW32367	Aaw32367 Mycobacte
14	1631	96.1	355	2 AAW32435	Aaw32435 Mycobacte
15	1631	96.1	355	2 AAW54307	Aaw54307 Mycobacte
16	1631	96.1	355	2 AAW81670	Aaw81670 M. tuberc
17	1631	96.1	355	2 AAY38972	Aay38972 M. tuberc
18	1631	96.1	355	2 AAY39109	Aay39109 M. tuberc
19	1631	96.1	355	4 AAU01890	Aau01890 M. tuberc
20	1631	96.1	355	5 AAE29701	Aae29701 Mycobacte
21	1631	96.1	355	5 AAE17565	Aae17565 Mycobacte
22	1000.5	58.9	1016	7 ADA26370	Ada26370 M. bovis
23	992	58.4	1154	7 ADA26368	Ada26368 Mycobacte
24	991.5	58.4	825	7 ADA26366	Ada26366 Mycobacte
25	990.5	58.3	813	7 ADA26367	Ada26367 Mycobacte

26	989	58.2	1022	7 ADA26369	Ada26369 Mycobacte
27	987	58.1	729	5 AAE29709	Aae29709 Mycobacte
28	987	58.1	729	5 AAE17573	Aae17573 Mycobacte
29	987	58.1	729	7 ADA26374	Ada26374 Mycobacte
30	984.5	58.0	875	7 ADA26365	Ada26365 Mycobacte
31	984	58.0	596	2 AAY32070	Aay32070 Mycobacte
32	984	58.0	596	5 AAE29710	Aae29710 Mycobacte
33	984	58.0	596	5 AAE17574	Aae17574 Mycobacte
34	984	58.0	599	5 AAU74599	Aau74599 Antigenic
35	984	58.0	729	4 AAO22142	Aao22142 Ra12-H9-3
36	984	58.0	729	5 AAE29708	Aae29708 Mycobacte
37	984	58.0	729	5 AAE17572	Aae17572 Mycobacte
38	984	58.0	729	7 ADA26373	Ada26373 Mycobacte
39	984	58.0	930	5 AAE29731	Aae29731 Mycobacte
40	984	58.0	930	5 ADA26364	Ada26364 Mycobacte
41	979	57.7	726	5 AAU74588	Aau74588 Antigenic
42	979	57.7	729	2 AAY32059	Aay32059 Mycobacte
43	978	57.6	195	5 AAE29704	Aae29704 Mycobacte
44	978	57.6	195	5 AAE17568	Aae17568 Mycobacte
45	676	39.8	231	2 AAY32071	Aay32071 Mycobacte

ALIGNMENTS

RESULT 1
AAE29703
ID AAE29703 standard; protein; 330 AA.
XX

AC AAE29703;
XX
DT 27-JAN-2003 (first entry)
XX
DE Mycobacterium tuberculosis Ra35FLMutSA mutant antigenic protein.
XX
KW Vaccine; immunity; diagnostic agent; gene therapy; Ra35FLMutSA; mutant;
antigen; mutein.
XX
OS Mycobacterium tuberculosis.
OS Synthetic.
FH Key Location/Qualifiers
FT Misc-difference 183
FT /note= "Wild type Ser substituted with Ala"
XX
PN WO200272792-A2.
XX
PD 19-SEP-2002.
XX
PF 13-MAR-2002; 2002WO-US008223.
XX
PR 13-MAR-2001; 2001US-0275837P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Brannon M, Guderian J;
XX
DR WPI: 2002-759844/82.
XX
DR N-PSDB; AAD47078.
XX
PT New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
against pathogenic microorganisms e.g. Leishmania and Mycobacterium
tuberculosis.
XX
PS Disclosure; Page 81-82; 155pp; English.
XX
CC The invention relates to a recombinant nucleic acid molecule encoding a
fusion polypeptide. The recombinant nucleic acid comprises a heterologous
polynucleotide sequence encoding an antigen or an antigenic fragment from
Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
polypeptide or its fragment. The Leishmania polynucleotide is selected
from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention

CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is M.
 CC tuberculosis Ra35FLMutSA mutant antigenic protein
 XX
 SQ Sequence 330 AA;

Query Match 100.0%; Score 1698; DB 5; Length 330;
 Best Local Similarity 100.0%; Pred. No. 1e-123;
 Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHHAPPALSDQRFADFPALPDPDSAMVAQVGPVNNINIKLGYNNVAVGAGTGIVID 60
 DB 1 MHHHHHAPPALSDQRFADFPALPDPDSAMVAQVGPVNNINIKLGYNNVAVGAGTGIVID 60
 QY 61 PNGVLTNNHVIAGATDINAFSVSGGQTYGVVVDYDRTQDVAVLQIRGAGLPSAAIGG 120
 DB 61 PNGVLTNNHVIAGATDINAFSVSGGQTYGVVVDYDRTQDVAVLQIRGAGLPSAAIGG 120
 QY 121 GVAVGEPVVMGNSGGGQTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQP 180
 DB 121 GVAVGEPVVMGNSGGGQTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQP 180
 QY 181 GDAGGPVVMGNSGGGQTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQP 240
 DB 181 GDAGGPVVMGNSGGGQTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQP 240
 QY 241 PTAFLGLGVDDNNGNGARVQVWGSAPASISLGTGDIITAVDGPINATAMADALNGH 300
 DB 241 PTAFLGLGVDDNNGNGARVQVWGSAPASISLGTGDIITAVDGPINATAMADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 2
 ID AAEL17567
 AC AAE17567 standard; protein; 330 AA.
 XX AAE17567;
 XX 22-APR-2002 (first entry)
 DT Mycobacterium species Ra35FL mature protein mutant (S183A), Ra35FLMutSA.
 DE Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; Ra32FLMutSA protein; mutant; mutin.
 XX Mycobacterium sp.

Key Location/Qualifiers
 FT Region 8..202
 FT /note= "Ra35 N-terminal peptide"
 FT Misc-difference 183
 FT /note= "Wild type Ser substituted with Ala"
 FT Region 199..330
 FT /note= "Ra35 C-terminal peptide, Ra12"

W0200198460-A2.
 XX
 XX 27-DEC-2001.
 XX 20-JUN-2001; 2001WO-US019959.
 XX 20-JUN-2000; 2000US-00597796.
 PR 01-FEB-2001; 2001US-0265737P.
 XX
 XX (CORI-) CORIXA CORP.

PI Skeiky Y, Reed S, Alderson M;
 XX WPI; 2002-147798/19.
 DR N-PSDB; AAD28337.
 XX
 PT Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 PS Claim 73; Fig 6; 136pp; English.

XX The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and the
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC Ra32FL mature protein mutant (S183A), Ra35FLMutSA

XX Sequence 330 AA;

Query Match 100.0%; Score 1698; DB 5; Length 330;
 Best Local Similarity 100.0%; Pred. No. 1e-123;
 Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSDQRFADFPALPDPDSAMVAQVGPVNNINIKLGYNNVAVGAGTGIVID 60
 DB 1 MHHHHHAPPALSDQRFADFPALPDPDSAMVAQVGPVNNINIKLGYNNVAVGAGTGIVID 60
 QY 61 PNGVLTNNHVIAGATDINAFSVSGGQTYGVVVDYDRTQDVAVLQIRGAGLPSAAIGG 120
 DB 61 PNGVLTNNHVIAGATDINAFSVSGGQTYGVVVDYDRTQDVAVLQIRGAGLPSAAIGG 120
 QY 121 GVAVGEPVVMGNSGGGQTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQP 180
 DB 121 GVAVGEPVVMGNSGGGQTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQP 180
 QY 181 GDAGGPVVMGNSGGGQTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQP 240
 DB 181 GDAGGPVVMGNSGGGQTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQP 240
 QY 241 PTAFLGLGVDDNNGNGARVQVWGSAPASISLGTGDIITAVDGPINATAMADALNGH 300
 DB 241 PTAFLGLGVDDNNGNGARVQVWGSAPASISLGTGDIITAVDGPINATAMADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 3
 ADA26372
 ID ADA26372 standard; protein; 330 AA.
 XX ADA26372;
 AC ADA26372;
 DT 20-NOV-2003 (first entry)
 XX
 DE Mycobacterium MTB32AMutSA protein.


```

Db      301 HPGDVISVTWTKSGGTRTGNVTLAEGPPA 330
RESULT 5
ADA26356
ID      ADA26356 standard; protein; 1010 AA.
XX
XX
AC      ADA26356;
XX
XX
DT      20-NOV-2003 (first entry)
XX
XX
DE      Mycobacterium MTB-102F fusion protein.
XX
XX
KW      fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW      tuberculosis; tuberculostatic; gene therapy; vaccine.
XX
XX
OS      Chimeric.
OS      Mycobacterium sp.
XX
XX
FN      WO2003070187-A2.
XX
XX
PD      28-AUG-2003.
XX
XX
XX      18-FEB-2003; 2003WO-US0004903.
XX
XX
PR      15-FEB-2002; 2002US-0357351P.
XX
XX
PA      (CORI-) CORIXA CORP.
XX
XX
XX      Skeiky Y, Guderian J, Reed S;
XX
XX
DR      WPI; 2003-697554/66.
DR      N-PSDB; ADA26355.
XX
XX
XX      New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT      and MTB39 antigens, with or without the MTB85A antigen, from a
PT      Mycobacterium species, useful for diagnosing, preventing and/or treating
PT      tuberculosis infection.
XX
XX
PS      Claim 6; Fig 4; 112pp; English.
XX
XX
XX      The invention relates to a novel nucleic acid encoding a fusion
CC      polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC      MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC      A polypeptide of the invention has tuberculostatic activity. A
CC      polynucleotide of the invention may have a use in gene therapy, and as a
CC      vaccine. The methods and compositions of the present invention are useful
CC      for diagnosing, preventing and/or treating tuberculosis infection. The
CC      present sequence is used in the exemplification of the invention.
XX
XX
SQ      Sequence 1010 AA;

Query Match      100.0%; Score 1698; DB 7; Length 1010;
Best Local Similarity 100.0%; Pred. No. 4,1e-123;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNAVAGAGTGIVID 60
Db      1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNAVAGAGTGIVID 60

Qy      61 PNGVLTNNHVIAGATDINAFSGSGGTGYDVVGVYDRTQDVAVLQIRAGGLPSAAIGG 120
Db      61 PNGVLTNNHVIAGATDINAFSGSGGTGYDVVGVYDRTQDVAVLQIRAGGLPSAAIGG 120

Qy      121 GVAVGEFVWAMGNSGGGGTPRAVPGRVALGQTVQASDSLTGAETLNLGIQFDAAIQP 180
Db      121 GVAVGEFVWAMGNSGGGGTPRAVPGRVALGQTVQASDSLTGAETLNLGIQFDAAIQP 180

Qy      181 GDAGGPPVNLGQVGVGMNTAASNLFQSQGGGFAIPIGQAMAIAQCIRSGGSPFVHIG 240
Db      181 GDAGGPPVNLGQVGVGMNTAASNLFQSQGGGFAIPIGQAMAIAQCIRSGGSPFVHIG 240

Qy      241 PTAFLGLGVDDNNGNGARVQRVWGSAPASLSLGISTGDTVITAVDGPAINSATAMADALNGH 300

```

```

Db      241 PTAFLGLGVDDNNGNGARVQRVWGSAPASLSLGISTGDTVITAVDGPAINSATAMADALNGH 300
Qy      301 HPGDVISVTWTKSGGTRTGNVTLAEGPPA 330
Db      301 HPGDVISVTWTKSGGTRTGNVTLAEGPPA 330

RESULT 6
AAE29702
ID      AAE29702 standard; protein; 330 AA.
XX
XX
AC      AAE29702;
XX
XX
DT      27-JAN-2003 (first entry)
XX
XX
DE      Mycobacterium tuberculosis mature Ra35 antigenic protein.
XX
XX
KW      Vaccine; immunity; diagnostic agent; gene therapy; Ra35 antigen.
XX
XX
OS      Mycobacterium tuberculosis.
XX
XX
FH      Key Location/Qualifiers
FT      Misc-difference 182 /note= "Encoded by GAG"
FT      Misc-difference 183 /note= "Encoded by GCG"
XX
XX
XX      WO200272792-A2.
XX
XX
PD      19-SEP-2002.
XX
XX
XX      13-MAR-2002; 2002WO-US008223.
XX
XX
PR      13-MAR-2001; 2001US-0275837P.
XX
XX
PA      (CORI-) CORIXA CORP.
XX
XX
XX      Skeiky Y, Brannon M, Guderian J;
XX
XX
DR      WPI; 2002-759844/82.
DR      N-PSDB; AAD47077.
XX
XX
XX      New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif,
PT      M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
PT      against pathogenic microorganisms e.g. Leishmania and Mycobacterium
PT      tuberculosis.
XX
XX
PS      Disclosure; Page 79-80; 155pp; English.
XX
XX
XX      The invention relates to a recombinant nucleic acid molecule encoding a
CC      fusion polypeptide. The recombinant nucleic acid comprises a heterologous
CC      polynucleotide sequence encoding an antigen or an antigenic fragment from
CC      Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
CC      polypeptide or its fragment. The Leishmania polynucleotide is selected
CC      from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention
CC      are used in methods for eliciting immune response in mammals. They are
CC      useful as vaccines to elicit protective immunity against pathogenic
CC      microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
CC      polypeptides are used for enhancing the expression of polynucleotides, as
CC      in vivo diagnostic agents and for raising antibodies in a non-human
CC      animal. The invention is used in gene therapy. The present sequence is M.
CC      tuberculosis mature Ra35 (N-terminus of MTB32A; RA35FL) antigenic protein
XX
XX
SQ      Sequence 330 AA;

Query Match      99.8%; Score 1695; DB 5; Length 330;
Best Local Similarity 99.7%; Pred. No. 1,7e-123;
Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNAVAGAGTGIVID 60
Db      1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINIKLGYNNAVAGAGTGIVID 60

```

QY 61 PNGVLTNNHVIAGATDINAFSGSGQTYGVVDVVGVDRTQDVAVLQIRGAGGLPSAAIGG 120
 DB 61 PNGVLTNNHVIAGATDINAFSGSGQTYGVVDVVGVDRTQDVAVLQIRGAGGLPSAAIGG 120
 QY 121 GVAVGEPVAMGNSGGGGTPRAVGRVWALGQTVQASDSLTGAEETLNGLIQFDAAIQP 180
 DB 121 GVAVGEPVAMGNSGGGGTPRAVGRVWALGQTVQASDSLTGAEETLNGLIQFDAAIQP 180
 QY 181 GDAGGPVNGLGQVVGNTAAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 DB 181 GDSGGPVNGLGQVVGNTAAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 QY 241 PTAFLGLGVDDNNGARVORVVGSAAPASISLGISTGDIVITAVDGAIPINSATADALNGH 300
 DB 241 PTAFLGLGVDDNNGARVORVVGSAAPASISLGISTGDIVITAVDGAIPINSATADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 7

AAE17566
 ID AAE17566 standard; protein; 330 AA.
 AC AAE17566;

22-APR-2002 (first entry)

Mycobacterium sp. MTB32A (Ra35FL) mature protein.

Fusion protein; antigen; serological sensitivity; immune response;
 tuberculosis; infection; vaccine; MTB32A; Ra32FL protein.

Mycobacterium sp.

Key Location/Qualifiers
 Region 8..202
 /note= "Ra35 N-terminal peptide"
 Misc-difference 182
 /note= "Encoded by GAG"
 Misc-difference 183
 /note= "Encoded by GCG"
 Region 199..330
 /note= "Ra35 C-terminal peptide, Ra12"

WO200198460-A2.

27-DEC-2001.

20-JUN-2001; 2001WO-US019959.

20-JUN-2000; 2000US-00597796.

01-FEB-2001; 2001US-0265737E.

(CORI-) CORIXA CORP.

Skeiky Y, Reed S, Alderson M;

WPI; 2002-147798/19.

N-PSDB; AAD28336.

Composition comprising MTB39 antigen and MTB32A antigen from
 Mycobacterium species, useful for eliciting immune response in a subject.

Claim 70; Fig 6; 136pp; English.

The present invention relates to fusion proteins containing at least two
 Mycobacterium species antigens, nucleotides encoding them and
 compositions comprising such fusion proteins. The present invention
 particularly relates to nucleic acids encoding fusion proteins that
 include two or more individual M. tuberculosis antigens which increase

CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB32A (Ra32FL) mature protein

XX Sequence 330 AA;

Query Match 99.8%; Score 1695; DB 5; Length 330;

Best Local Similarity 99.7%; Fred. No. 1.7e-123;

Matches 329; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPVVNINIKLGYNNAVGAGTGIVID 60
 DB 1 MHHHHHAPPALSQDRFADFPALPLDPSAMVAQVGPVVNINIKLGYNNAVGAGTGIVID 60
 QY 61 PNGVLTNNHVIAGATDINAFSGSGQTYGVVDVVGVDRTQDVAVLQIRGAGGLPSAAIGG 120
 DB 61 PNGVLTNNHVIAGATDINAFSGSGQTYGVVDVVGVDRTQDVAVLQIRGAGGLPSAAIGG 120
 QY 121 GVAVGEPVAMGNSGGGGTPRAVGRVWALGQTVQASDSLTGAEETLNGLIQFDAAIQP 180
 DB 121 GVAVGEPVAMGNSGGGGTPRAVGRVWALGQTVQASDSLTGAEETLNGLIQFDAAIQP 180
 QY 181 GDAGGPVNGLGQVVGNTAAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 DB 181 GDSGGPVNGLGQVVGNTAAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIG 240
 QY 241 PTAFLGLGVDDNNGARVORVVGSAAPASISLGISTGDIVITAVDGAIPINSATADALNGH 300
 DB 241 PTAFLGLGVDDNNGARVORVVGSAAPASISLGISTGDIVITAVDGAIPINSATADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330

RESULT 8

ADA26371

ID ADA26371 standard; protein; 330 AA.

AC ADA26371;

DT 20-NOV-2003 (first entry)

DE Mycobacterium wild type MBT32A protein.

KW fusion polypeptide; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculostatic; gene therapy; vaccine.

OS Chimeric.

OS Mycobacterium sp.

PN WO2003070187-A2.

PD 28-AUG-2003.

PF 18-FEB-2003; 2003WO-US004903.

PR 15-FEB-2002; 2002US-0357351P.

PA (CORI-) CORIXA CORP.

XX PI Skeiky Y, Guderian J, Reed S;
 XX DR WPI; 2003-697554/66.
 XX
 XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 PT and MTB39 antigens, with or without the MTB85A antigen, from a
 PT Mycobacterium species, useful for diagnosing, preventing and/or treating
 PT tuberculosis infection.
 XX
 XX Disclosure; Fig 19; 112pp; English.
 PS
 CC The invention relates to a novel nucleic acid encoding a fusion
 CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
 CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
 CC A polypeptide of the invention has tuberculostatic activity. A
 CC polynucleotide of the invention may have a use in gene therapy, and as a
 CC vaccine. The methods and compositions of the present invention are useful
 CC for diagnosing, preventing and/or treating tuberculosis infection. The
 CC present sequence is used in the exemplification of the invention.
 XX
 XX Sequence 330 AA;
 SQ
 Query Match 99.4%; Score 1688; DB 7; Length 330;
 Best Local Similarity 99.4%; Pred. No. 6.1e-123;
 Matches 328; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MHHHHHAPPALSODRFADFPALPLDPSAMVAQVGPVNNVNTKLGYNNAVAGTGTGIVID 60
 DB 1 MHHHHHAPPALSODRFADFPALPLDPSAMVAQVGPVNNVNTKLGYNNAVAGTGTGIVID 60
 QY 61 PNGVLTNNHVIAGATDINAFSGSGGTGYGVVGYDRTQDVAVLQLRGAGGLPSAAIGG 120
 DB 61 PNGVLTNNHVIAGATDINAFSGSGGTGYGVVGYDRTQDVAVLQLRGAGGLPSAAIGG 120
 QY 121 GVAVGEPVWAMNSGGGGTTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
 DB 121 GVAVGEPVWAMNSGGGGTTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQP 180
 QY 181 GDAGGPPVNLGQVVGMMNTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSGPTVHIG 240
 DB 181 GDAGGPPVNLGQVVGMMNTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSGPTVHIG 240
 QY 241 PTAFLGLGVNDNNGARVQVWGSAAPASLGISTGDIVITAVDGPINSATAMADALNGH 300
 DB 241 PTAFLGLGVNDNNGARVQVWGSAAPASLGISTGDIVITAVDGPINSATAMADALNGH 300
 QY 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 DB 301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA 330
 RESULT 9
 ID AAY05000 standard; protein; 355 AA.
 XX
 XX AAY05000;
 AC
 DT 06-JUL-1999 (first entry)
 XX
 XX Mycobacterium species protein sequence 50D.
 DE
 DE Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.
 XX
 XX Mycobacterium sp.
 OS
 XX WO9909186-A2.
 PN
 PD 25-FEB-1999.
 XX
 XX 14-AUG-1998; 98WO-FR001813.
 PF
 XX

PR 14-AUG-1997; 97FR-00010404.
 PR 11-SEP-1997; 97FR-00011325.
 XX
 XX (INSP) INST PASTEUR.
 PA
 XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
 PI Goguet De La Salmoniere Y;
 XX
 XX WPI; 1999-181045/15.
 DR N-PSDB; AAX34251.
 XX
 XX Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in infection-associated
 PT protein expression.
 PT
 XX Claim 32; Fig 50D; 309pp; French.
 PS
 XX Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
 CC from various Mycobacterium species microorganisms. The encoding
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection
 CC
 XX Sequence 355 AA;
 SQ
 Query Match 96.7%; Score 1642; DB 2; Length 355;
 Best Local Similarity 99.7%; Pred. No. 2.5e-119;
 Matches 322; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVNNVNTKLGYNNAVAGTGTGIVIDPVGWVLT 67
 DB 33 APPALSQDRFADFPALPLDPSAMVAQVGPVNNVNTKLGYNNAVAGTGTGIVIDPVGWVLT 92
 QY 68 NNHVIAGATDINAFSGSGGTGYGVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
 DB 93 NNHVIAGATDINAFSGSGGTGYGVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
 QY 128 VVAMGNSGGGGTTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPDAGGPV 187
 DB 153 VVAMGNSGGGGTTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPDAGGPV 212
 QY 188 VNGLGQVVGMMNTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSGPTVHIGPTAFGL 247
 DB 213 VNGLGQVVGMMNTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSGPTVHIGPTAFGL 272
 QY 248 GVVDNNGARVQVWGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS 307
 DB 273 GVVDNNGARVQVWGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVIS 332
 QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
 DB 333 VTWQTKSGGTRTGNVTLAEGPPA 355
 RESULT 10
 ID AAO22137
 XX
 XX AAO22137 standard; protein; 355 AA.
 AC
 DT 03-OCT-2002 (first entry)
 XX
 XX Mycobacterium tuberculosis MTB32A protein.
 DE
 DE Ral2; serine protease antigen; MTB32A; Mycobacterium tuberculosis;
 KW vaccine; eukaryotic; prokaryotic origin; serum antibody; fusion protein;
 KW immunogen; cytokine.
 XX
 XX Mycobacterium tuberculosis.
 OS
 XX WO200125401-A2.
 PN
 XX

Db 333 VTWTKSGGTRTGNVTLAEGPPA 355

RESULT 12

AA04830
ID AAY04830 standard; protein; 379 AA.

XX AC AAY04830;

XX DT 06-JUL-1999 (first entry)

XX XX Mycobacterium species protein sequence 50F.

XX DE Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
XX KW hybridisation; detection; vaccine; immunisation; infection.

XX XX Mycobacterium sp.

XX OS WO9909186-A2.

XX FN 25-FEB-1999.

XX PD 14-AUG-1998; 98WO-FR001813.

XX PF 14-AUG-1997; 97FR-00010404.

XX PR 11-SEP-1997; 97FR-00011325.

XX XX (INSP) INST PASTEUR.

XX PA Gicquel B, Portnoie D, Lim B, Pelicic V, Guigueno A;

XX PI Goguet De La Salmoniere Y;

XX FI WPI; 1999-181045/15.

XX DR N-PSDB; AAX34252.

XX PT Mycobacterial DNA vectors containing reporter constructs - for

XX PT identifying coding or promoter sequences involved in infection-associated

XX PT protein expression.

XX PS Claim 32; Fig 50F; 309pp; French.

XX XX Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins

XX CC from various Mycobacterium species microorganisms. The encoding

XX CC nucleotide sequences can be used as primers and probes for methods for

XX CC detecting and identifying mycobacteria, especially belonging to the M.

XX CC tuberculosis complex. The encoded proteins can be used in vaccines for

XX CC immunisation against a bacterial or viral infection

XX SQ Sequence 379 AA;

Query Match 96.7%; Score 1642; DB 2; Length 379;

Best Local Similarity 99.7%; Pred. No. 2.7e-119;

Matches 322; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 8 APPALSQDRFADFPALPLDPSAMVAQVQVNTKLGYNNAVAGTGVIDPNGVLT 67

Db 57 APPALSQDRFADFPALPLDPSAMVAQVQVNTKLGYNNAVAGTGVIDPNGVLT 116

Qy 68 NNHVIAGATDINAFSVGSGQTYGVVDVGYDRTQDVAVLQLRGAGLPSAALGGGVAVGEP 127

Db 117 NNHVIAGATDINAFSVGSGQTYGVVDVGYDRTQDVAVLQLRGAGLPSAALGGGVAVGEP 176

Qy 128 VVAMGNSGGQGTTPRAVPGRVVAGLQGVVQASDSLTGAETLNGLIQFDAAIQPDGAGGPV 187

Db 177 VVAMGNSGGQGTTPRAVPGRVVAGLQGVVQASDSLTGAETLNGLIQFDAAIQPDGAGGPV 236

Qy 188 VNLGQVVGMMNTAASDNFQLSQGGQGFAPICQMAIAGQIRSGGSPVTHIGTAFGL 247

Db 237 VNLGQVVGMMNTAASDNFQLSQGGQGFAPICQMAIAGQIRSGGSPVTHIGTAFGL 296

Qy 248 GVVDNNGNGARVQVVGSAASLGISTGVITAVDGAFINSATAMADALNGHHPGDVIS 307

Db 297 GVVDNNGNGARVQVVGSAASLGISTGVITAVDGAFINSATAMADALNGHHPGDVIS 356

Qy 308 VTWTKSGGTRTGNVTLAEGPPA 330

Db 357 VTWTKSGGTRTGNVTLAEGPPA 379

RESULT 13

AAW32367

ID AAW32367 standard; protein; 355 AA.

XX AC AAW32367;

XX DT 13-JAN-1998 (first entry)

XX DE Mycobacterium tuberculosis antigen TbrA35.

XX KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

XX OS Mycobacterium tuberculosis.

XX FN WO9709429-A2.

XX PD 13-MAR-1997.

XX PF 30-AUG-1996; 96WO-US014675.

XX PR 01-SEP-1995; 95US-00523435.

XX PR 22-SEP-1995; 95US-00532136.

XX PR 22-MAR-1996; 96US-00620280.

XX PR 05-JUN-1996; 96US-00658800.

XX PR 12-JUL-1996; 96US-00680573.

XX PA (CORI-) CORIXA CORP.

XX PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;

XX PI Vedwick TH, Twardzik DR;

XX DR WPI; 1997-192904/17.

XX DR N-PSDB; AAT91414.

XX PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -

XX PT useful for diagnosis of M. tuberculosis infection.

XX PS Example 3; Page 124-126; 190pp; English.

XX CC A new immunogenic polypeptide has been developed comprising an

XX CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its

XX CC variant differing only in conservative substitutions and/or

XX CC modifications). The present sequence represents a M.tuberculosis antigen,

XX CC TbrA35. The immunogenic polypeptide can be used to diagnose

XX CC M.tuberculosis infection by forming complexes with specific antibodies in

XX CC the sample. Fragments of DNA encoding the immunogenic polypeptide can be

XX CC used as diagnostic primers or probes and agents that bind to the antigen,

XX CC especially monoclonal antibodies or equivalent polyclonal antibodies, are

XX CC also used for diagnosis

XX SQ Sequence 355 AA;

Query Match 96.1%; Score 1631; DB 2; Length 355;

Best Local Similarity 99.1%; Pred. No. 1.8e-118;

Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 8 APPALSQDRFADFPALPLDPSAMVAQVQVNTKLGYNNAVAGTGVIDPNGVLT 67

Db 33 APPALSQDRFADFPALPLDPSAMVAQVQVNTKLGYNNAVAGTGVIDPNGVLT 92

Qy 68 NNHVIAGATDINAFSVGSGQTYGVVDVGYDRTQDVAVLQLRGAGLPSAALGGGVAVGEP 127

Db 93 NNHVIAGATDINAFSVGSGQTYGVVDVGYDRTQDVAVLQLRGAGLPSAALGGGVAVGEP 152

Qy 128 VVAMGNSGGQGTTPRAVPGRVVAGLQGVVQASDSLTGAETLNGLIQFDAAIQPDGAGGPV 187

Db 153 VVAMNGSGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQPDGSGPV 212
 Qy 188 VNLGQVVGMNTAASDNFQLSQGGQFAIPICQAMAIAGQIRSGGSPVTHIGTAFGL 247
 Db 213 VNLGQVVGMNTAASDNFQLSQGGQFAIPICQAMAIAGQIRSGGSPVTHIGTAFGL 272
 Qy 248 GVVDNNGGARVORVVGSAAPASLIGISTGDTVITAVDGPINSATAMADALNGHHPGDVIS 307
 Db 273 GVVDNNGGARVORVVGSAAPASLIGISTGDTVITAVDGPINSATAMADALNGHHPGDVIS 332
 Qy 308 VTWQTKSGGTRTGNVTLAEGPPA 330
 Db 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 14
 AAW32435
 ID AAW32435 standard; protein; 355 AA.
 XX
 AC AAW32435;
 XX
 DT 08-JAN-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen Tbra35.
 XX
 KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 KW skin testing; M.tuberculosis.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9709428-A2.
 XX
 PD 13-MAR-1997.
 XX
 PF 30-AUG-1996; 96WO-US014674.
 XX
 PR 01-SEP-1995; 95US-00523436.
 PR 22-SEP-1995; 95US-00533634.
 PR 22-MAR-1996; 96US-00620874.
 PR 05-JUN-1996; 96US-00659683.
 PR 12-JUL-1996; 96US-00680574.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Read SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TH, Twardzik DR;
 XX
 XX WPI; 1997-192903/17.
 DR N-PSDB; AAT91477.
 XX
 XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also for
 PT diagnosis.
 XX
 PS Example 3; Page 114-116; 168pp; English.
 XX
 CC A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
 CC variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis antigen,
 CC Tbra35. The immunogenic protein, and fusion proteins containing one or
 CC more of the proteins or one of the proteins plus ESAT-6, are useful in
 CC vaccines, preferably when formulated with a non-specific adjuvant, to
 CC induce an immune response against M.tuberculosis (for treatment or
 CC prevention)
 XX
 XX Sequence 355 AA;
 Query Match 96.1%; Score 1631; DB 2; Length 355;
 Best Local Similarity 99.1%; Pred. No. 1.8e-118;
 Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 8 APPALSQDRFADFPALPLDPSAMVAQVQVNVNINIKLGYNNAGAGTIVIDPNGVLT 67

Db 33 APPALSQDRFADFPALPLDPSAMVAQVQVNVNINIKLGYNNAGAGTIVIDPNGVLT 92
 Qy 68 NNHVIAGATDINAFVSGGQTVGVVDVGYDRTQDVAVIQLRGAGGLPSAAICGGVAVGEP 127
 Db 93 NNHVIAGATDINAFVSGGQTVGVVDVGYDRTQDVAVIQLRGAGGLPSAAICGGVAVGEP 152
 Qy 128 VVAMNGSGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQPDGSGPV 187
 Db 153 VVAMNGSGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQPDGSGPV 212
 Qy 188 VNLGQVVGMNTAASDNFQLSQGGQFAIPICQAMAIAGQIRSGGSPVTHIGTAFGL 247
 Db 213 VNLGQVVGMNTAASDNFQLSQGGQFAIPICQAMAIAGQIRSGGSPVTHIGTAFGL 272
 Qy 248 GVVDNNGGARVORVVGSAAPASLIGISTGDTVITAVDGPINSATAMADALNGHHPGDVIS 307
 Db 273 GVVDNNGGARVORVVGSAAPASLIGISTGDTVITAVDGPINSATAMADALNGHHPGDVIS 332
 Qy 308 VTWQTKSGGTRTGNVTLAEGPPA 330
 Db 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 15
 AAW64307
 ID AAW64307 standard; protein; 355 AA.
 XX
 AC AAW64307;
 XX
 DT 17-OCT-2003 (revised)
 DT 09-NOV-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen Tbra35.
 XX
 KW Tuberculosis; infection; diagnosis; antigen; Tbra35.
 XX
 OS Mycobacterium tuberculosis; strain H37Ra.
 XX
 PN WO9816645-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US018214.
 XX
 PR 11-OCT-1996; 96US-00729622.
 PR 13-MAR-1997; 97US-00818111.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Read SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX
 XX WPI; 1998-251292/22.
 DR N-PSDB; AAW44355.
 XX
 XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and
 PT diagnosis of tuberculosis.
 XX
 PS Example 3; Page 115-116; 250pp; English.
 XX
 CC This polypeptide comprises Mycobacterium tuberculosis soluble antigen
 CC Tbra35. It is encoded by a DNA sequence (see AAW44355) isolated from a M.
 CC tuberculosis strain H37Ra expression library with rabbit anti-sera raised
 CC against M. tuberculosis supernatant. No significant homology was found
 CC between Tbra35 and Genbank database sequences. The invention relates to
 CC compositions and methods for diagnosing tuberculosis. It provides
 CC polypeptides (see AAW64291-W64373) comprising an antigenic portion of a
 CC soluble M. tuberculosis antigen, or an immunogenic portion of an M.
 CC tuberculosis antigen, as well as DNA sequences encoding such
 CC polypeptides, recombinant expression vectors and transformed or
 CC transfected host cells. Also claimed are methods and diagnostic kits for

Wed Jun 23 16:34:40 2004

CC detecting M. tuberculosis infection in a patient using the above
CC polypeptides, antibodies or oligonucleotide probes and primers, for the
CC diagnosis of tuberculosis. (Updated on 17-OCT-2003 to standardise OS
CC field)
XX
SQ Sequence 355 AA;
Query Match 96.1%; Score 1631; DB 2; Length 355;
Best Local Similarity 99.1%; Pred. No. 1.8e-118;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 8 APPALSDRFPALPLDPSAMVAQVPOVNVNINIKLGYNNAVGAGTGIVIDPNGWVLT 67
Db 33 APPALSDRFPALPLDPSAMVAQVPOVNVNINIKLGYNNAVGAGTGIVIDPNGVLT 92
Qy 68 NNHVIAGATDINAFSVSGSGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
Db 93 NNHVIAGATDINAFSVSGSGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
Qy 128 VVAMGNSGGGGTTPRAVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPDAGGPV 187
Db 153 VVAMGNSGGGGTTPRAVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPDSGGPV 212
Qy 188 VNLGQVVGMNTAASDNFQSQGGQGFAPIGQAMAIAGQIRSGGSGPTVHIGTFAFLGL 247
Db 213 VNLGQVVGMNTAASDNFQSQGGQGFAPIGQAMAIAGQIRSGGSGPTVHIGTFAFLGL 272
Qy 248 GVDNNGNGARVQRVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 307
Db 273 GVDNNGNGARVQRVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGHHPGDVIS 332
Qy 308 VTWQTKSGGTRTGNVTLAEGPPA 330
Db 333 VNWQTKSGGTRTGNVTLAEGPPA 355

Search completed: June 22, 2004, 17:15:25
Job time : 41.5137 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:07:54 ; Search time 11.78 Seconds
(without alignments)
1446.225 Million cell updates/sec

Title: US-09-886-349A-6

Perfect score: 1698

Sequence: 1 MHHHHHAPPALSQDRFADF.....QTKSGGTGTCNTLAEGPPA 330

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*

2: /cgn2_6/prodata/2/iaa/5B COMB.pep.*

3: /cgn2_6/prodata/2/iaa/6A COMB.pep.*

4: /cgn2_6/prodata/2/iaa/6B COMB.pep.*

5: /cgn2_6/prodata/2/iaa/PCUTUS COMB.pep.*

6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1631	96.1	355	3	US-08-818-112-79
2	1631	96.1	355	4	US-08-818-111-80
3	1631	96.1	355	4	US-09-556-556-79
4	1631	96.1	355	4	US-09-072-596-80
5	1631	96.1	355	4	US-09-072-967-79
6	984	58.0	596	4	US-09-287-849-26
7	984	58.0	729	4	US-09-223-040-2
8	984	58.0	729	4	US-09-287-849-2
9	676	39.8	231	4	US-09-287-849-28
10	670	39.5	132	3	US-08-818-112-66
11	670	39.5	132	4	US-08-818-111-67
12	670	39.5	132	4	US-09-556-556-66
13	670	39.5	132	4	US-09-072-596-67
14	670	39.5	132	4	US-09-072-967-66
15	670	39.5	132	4	US-09-636-215-819
16	670	39.5	132	4	US-09-636-215-848
17	670	39.5	132	4	US-09-585-166A-819
18	670	39.5	132	4	US-09-585-166A-848
19	633	37.3	224	4	US-09-636-215-825
20	633	37.3	224	4	US-09-685-166A-825
21	633	37.3	267	4	US-09-643-597-352
22	633	37.3	267	4	US-09-606-421B-352
23	633	37.3	273	4	US-09-736-457-1864
24	633	37.3	299	4	US-09-643-597-354
25	633	37.3	299	4	US-09-606-421B-354
26	633	37.3	304	4	US-09-636-215-835
27	633	37.3	304	4	US-09-685-166A-835

28	633	37.3	314	4	US-09-736-457-1863	Sequence 1863, App
29	633	37.3	400	4	US-09-636-215-852	Sequence 852, App
30	633	37.3	400	4	US-09-685-166A-852	Sequence 852, App
31	633	37.3	487	4	US-09-620-412C-349	Sequence 349, App
32	633	37.3	487	4	US-09-598-419-349	Sequence 349, App
33	633	37.3	518	4	US-09-620-412C-333	Sequence 333, App
34	633	37.3	518	4	US-09-598-419-333	Sequence 333, App
35	633	37.3	525	4	US-09-556-877-196	Sequence 196, App
36	633	37.3	525	4	US-09-620-412C-196	Sequence 196, App
37	633	37.3	525	4	US-09-598-419-196	Sequence 196, App
38	633	37.3	583	4	US-09-620-412C-353	Sequence 353, App
39	633	37.3	583	4	US-09-598-419-353	Sequence 353, App
40	633	37.3	585	4	US-09-620-412C-337	Sequence 337, App
41	633	37.3	585	4	US-09-598-419-337	Sequence 337, App
42	633	37.3	619	4	US-09-620-412C-309	Sequence 309, App
43	633	37.3	619	4	US-03-598-419-309	Sequence 309, App
44	633	37.3	631	4	US-09-620-412C-325	Sequence 325, App
45	633	37.3	631	4	US-09-598-419-325	Sequence 325, App

ALIGNMENTS

RESULT 1
US-08-818-112-79
; Sequence 79, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-79

Query Match 96.1%; Score 1631; DB 3; Length 355;
Best Local Similarity 99.1%; Pred. No. 3.9e-134;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNTKLGYNNAVAGTGIIVDPNGVLT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVAPQVNTKLGYNNAVAGTGIIVDPNGVLT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVYVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVYVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
QY 128 VVAMNSGGGGTTPRAYGPRVVALQQTVAQSDSLTGAEETLNGLIQFDDAAIQPDGAGGPV 187
DB 153 VVAMNSGGGGTTPRAYGPRVVALQQTVAQSDSLTGAEETLNGLIQFDDAAIQPDGAGGPV 212
QY 188 VVNGLGQVVMNTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPVTHIGPTAFLGL 247
DB 213 VVNGLGQVVMNTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPVTHIGPTAFLGL 272
QY 248 GVVDNNGNGARVQVVGSAAPASLGISTGDIITAVDGPINSATAMADALNGHHPGDVIS 307
DB 273 GVVDNNGNGARVQVVGSAAPASLGISTGDIITAVDGPINSATAMADALNGHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 2
US-08-818-111-80
; Sequence 80, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-80
Query Match 96.1%; Score 1631; DB 4; Length 355;
Best Local Similarity 99.1%; Pred. No. 3.9e-134;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNTKLGYNNAVAGTGIIVDPNGVLT 67
DB 33 APPALSQDRFADFPALPLDPSAMVAQVAPQVNTKLGYNNAVAGTGIIVDPNGVLT 92
QY 68 NNHVIAGATDINAFSVSGGQTYGVYVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
DB 93 NNHVIAGATDINAFSVSGGQTYGVYVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
QY 128 VVAMNSGGGGTTPRAYGPRVVALQQTVAQSDSLTGAEETLNGLIQFDDAAIQPDGAGGPV 187
DB 153 VVAMNSGGGGTTPRAYGPRVVALQQTVAQSDSLTGAEETLNGLIQFDDAAIQPDGAGGPV 212
QY 188 VVNGLGQVVMNTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPVTHIGPTAFLGL 247
DB 213 VVNGLGQVVMNTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPVTHIGPTAFLGL 272
QY 248 GVVDNNGNGARVQVVGSAAPASLGISTGDIITAVDGPINSATAMADALNGHHPGDVIS 307
DB 273 GVVDNNGNGARVQVVGSAAPASLGISTGDIITAVDGPINSATAMADALNGHHPGDVIS 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
DB 333 VNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 3
US-09-056-556-79
; Sequence 79, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-79
Query Match 96.1%; Score 1631; DB 4; Length 355;
Best Local Similarity 99.1%; Pred. No. 3.9e-134;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSQDRFADFPALPLDPSAMVAQVGPVQVNTKLGYNNAVAGTGIIVDPNGVLT 67

Db 33 APPALSDQRFADFPALPLDPSAMVAQVAVQVNNINTKLGYNNNAVAGTGIVIDPNGVVL 92
QY 68 NNHVIAGATDINAFSVGSGGTQYVDVVGVDRTQDVAVLQIRGAGGLPSAAICGGVAVGEP 127
Db 93 NNHVIAGATDINAFSVGSGGTQYVDVVGVDRTQDVAVLQIRGAGGLPSAAICGGVAVGEP 152
QY 128 VVAMGNSGGGGTTPRAVPRVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPV 187
Db 153 VVAMGNSGGGGTTPRAVPRVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPV 212
QY 188 VVAMGNSGGGGTTPRAVPRVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPV 247
Db 213 VVAMGNSGGGGTTPRAVPRVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPV 272
QY 248 GVVDNNGNGARVQRVVGSAAPASLIGTGDVITAVDGPINSATAMADALNGHHPGDV 307
Db 273 GVVDNNGNGARVQRVVGSAAPASLIGTGDVITAVDGPINSATAMADALNGHHPGDV 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
Db 333 VVWQTKSGGTRTGNVTLAEGPPA 355

RESULT 4
US-09-072-596-80
; Sequence 80, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-80

Query Match 96.1%; Score 1631; DB 4; Length 355;
Best Local Similarity 99.1%; Pred. No. 3.9e-134;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 APPALSDQRFADFPALPLDPSAMVAQVAVQVNNINTKLGYNNNAVAGTGIVIDPNGVVL 67
Db 33 APPALSDQRFADFPALPLDPSAMVAQVAVQVNNINTKLGYNNNAVAGTGIVIDPNGVVL 92
QY 68 NNHVIAGATDINAFSVGSGGTQYVDVVGVDRTQDVAVLQIRGAGGLPSAAICGGVAVGEP 127
Db 93 NNHVIAGATDINAFSVGSGGTQYVDVVGVDRTQDVAVLQIRGAGGLPSAAICGGVAVGEP 152
QY 128 VVAMGNSGGGGTTPRAVPRVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPV 187
Db 153 VVAMGNSGGGGTTPRAVPRVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPV 212
QY 188 VVAMGNSGGGGTTPRAVPRVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPV 247
Db 213 VVAMGNSGGGGTTPRAVPRVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDAGGPV 272
QY 248 GVVDNNGNGARVQRVVGSAAPASLIGTGDVITAVDGPINSATAMADALNGHHPGDV 307
Db 273 GVVDNNGNGARVQRVVGSAAPASLIGTGDVITAVDGPINSATAMADALNGHHPGDV 332
QY 308 VTWQTKSGGTRTGNVTLAEGPPA 330
Db 333 VVWQTKSGGTRTGNVTLAEGPPA 355

RESULT 5
US-09-072-967-79
; Sequence 79, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-967-79

Query Match 96.1%; Score 1631; DB 4; Length 355;
Best Local Similarity 99.1%; Pred. No. 3.9e-134;
Matches 320; Conservative 1; Mismatches 2; Indels

QY	8	APPALSDRFDADFPALPLDPSAMTAQVGPVNNINTKLYNNNAVAGAGTGIVTDPNGVLT	67
DB	33	APPALSDRFDADFPALPLDPSAMTAQVGPVNNINTKLYNNNAVAGAGTGIVTDPNGVLT	92
QY	68	NNHVAGATDINAFSVSGGQTYGYDVVGYDRTQDVAVLQRGAGSLPSAAIGGGVAVGEP	127
DB	93	NNHVAGATDINAFSVSGGQTYGYDVVGYDRTQDVAVLQRGAGSLPSAAIGGGVAVGEP	152
QY	128	VYAMGNSGGCGTTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIOPGDSAGPV	187
DB	153	VYAMGNSGGCGTTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIOPGDSAGPV	212
QY	188	VNGLGQVYVMNTAASDNFQLSGGGGFAIPIGAMAIAGQIRSGGSPVTHIGTAFJGL	247
DB	213	VNGLGQVYVMNTAASDNFQLSGGGGFAIPIGAMAIAGQIRSGGSPVTHIGTAFJGL	272
QY	248	GYVDNNGNGARVQVVGSAPAASIGTSTGVITAVDGPINSATAMADALNHGHPGDVIS	307
DB	273	GYVDNNGNGARVQVVGSAPAASIGTSTGVITAVDGPINSATAMADALNHGHPGDVIS	332
QY	308	VTWQTKSGGTRTGNVTLAEGPPA	330
DB	333	VNWQTKSGGTRTGNVTLAEGPPA	355

```

RESULT 6
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Nelo, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1993-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: bi-fusion
US-09-287-849-26

```

[illegible]

```

RESULT 7
US-09-223-040-2
; Sequence 2, Application US/09223040
; Patent No. 6544522
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: C14058-009010US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
; US-09-223-040-2

```

```

RESULT 8
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US

```



```
/ CURRENT APPLICATION NUMBER: US/09/287,849
/ CURRENT FILING DATE: 1999-04-07
/ PRIOR APPLICATION NUMBER: US 08/818,112
/ PRIOR FILING DATE: 1997-03-13
/ PRIOR APPLICATION NUMBER: US 08/942,578
/ PRIOR FILING DATE: 1997-10-01
/ PRIOR APPLICATION NUMBER: US 09/025,197
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 09/056,556
/ PRIOR FILING DATE: 1998-04-07
/ PRIOR APPLICATION NUMBER: US 09/223,040
/ PRIOR FILING DATE: 1998-12-30
/ NUMBER OF SEQ ID NOS: 46
/ SOFTWARE: Patent in Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 729
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-2

Query Match          58.0%; Score 984; DB 4; Length 729;
Best Local Similarity 99.5%; Pred. No. 2.2e-77;
Matches 194; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      8 APPALSQDRFADPPALPLDPSAMVAGQGVVNNINTKLGYNNVAGAGTGIVIDPNGVLT 67
Db      535 APPALSQDRFADPPALPLDPSAMVAGQGVVNNINTKLGYNNVAGAGTGIVIDPNGVLT 594
QY      68 NNHVIAGATDINAFSGSGTGVGVVGYDRDQDVAVLQLRGAGGLPSAAIGGVAVGEP 127
Db      595 NNHVIAGATDINAFSGSGTGVGVVGYDRDQDVAVLQLRGAGGLPSAAIGGVAVGEP 654
QY      128 VVAMNGSGGGTTPRAVGRVVALGTVQASDLSLTGAETLNLQIFDAAIQPGDAGGPV 187
Db      655 VVAMNGSGGGTTPRAVGRVVALGTVQASDLSLTGAETLNLQIFDAAIQPGDAGGPV 714
QY      188 VNLGQGVVGNNTAAS 202
Db      715 VNLGQGVVGNNTAAS 729

RESULT 9
US-09-287-849-28
/ Sequence 28, Application US/09287849
/ Patent No. 6627198
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Alderson, Mark
/ APPLICANT: Campos-Neto, Antonio
/ APPLICANT: Corixa Corporation
/ TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
/ TITLE OF INVENTION: And Their Uses
/ FILE REFERENCE: 014058-009020US
/ CURRENT APPLICATION NUMBER: US/09/287,849
/ CURRENT FILING DATE: 1999-04-07
/ PRIOR APPLICATION NUMBER: US 08/818,112
/ PRIOR FILING DATE: 1997-03-13
/ PRIOR APPLICATION NUMBER: US 08/942,578
/ PRIOR FILING DATE: 1997-10-01
/ PRIOR APPLICATION NUMBER: US 09/025,197
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 09/056,556
/ PRIOR FILING DATE: 1998-04-07
/ PRIOR APPLICATION NUMBER: US 09/223,040
/ PRIOR FILING DATE: 1998-12-30
/ NUMBER OF SEQ ID NOS: 46
/ SOFTWARE: Patent in Ver. 2.1
/ SEQ ID NO 28
/ LENGTH: 231

/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: bi-fusion
US-09-287-849-28

Query Match          39.8%; Score 676; DB 4; Length 231;
Best Local Similarity 99.2%; Pred. No. 3e-51;
Matches 132; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      198 NTAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPTVHIGPTAFILGLGVVDNNGGA 257
Db      8 HTAASDNFQLSQGGGFAIPIGQAWAIAQIRSGGSPTVHIGPTAFILGLGVVDNNGGA 67
QY      258 RVQRVWGSAPAAASLGISTGVDVITAVDGAPINSATAMADALNHHHPGDDVISVTWQTKSGT 317
Db      68 RVQRVWGSAPAAASLGISTGVDVITAVDGAPINSATAMADALNHHHPGDDVISVTWQTKSGT 127
QY      318 RTGNVTTLAEGPPA 330
Db      128 RTGNVTTLAEGPPA 140

RESULT 10
US-08-818-112-66
/ Sequence 66, Application US/08818112
/ Patent No. 6290969
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Dillon, Davin C.
/ APPLICANT: Campos-Neto, Antonio
/ APPLICANT: Houghton, Raymond
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Twardzik, Daniel R.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
/ TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
/ NUMBER OF SEQUENCES: 153
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SEED AND BERRY LLP
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: USA
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/818,112
/ FILING DATE: 13-MAR-1997
/ CLASSIFICATION: 424
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Maki, David J.
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.411C6
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 66:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 132 amino acids
/ TYPE: amiro acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
US-08-818-112-66

Query Match          39.5%; Score 670; DB 3; Length 132;
Best Local Similarity 99.2%; Pred. No. 4.6e-51;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 199 TAASDNFQLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 258
 Db 1 TAASDNFQLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
 QY 259 VORVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVIVSVTWOTKSGGTR 318
 Db 61 VORVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVIVSVTWOTKSGGTR 120
 QY 319 TGNVTLAEGPPA 330
 Db 121 TGNVTLAEGPPA 132

RESULT 11
 US-08-818-111-67
 ; Sequence 67, Application US/08818111
 ; Patent No. 6338852
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonia
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Twardzik, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 148
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/818,111
 ; FILING DATE: 13-MAR-1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 210121.417C6
 ; REFERENCE/DOCKET NUMBER: 210121.417C6
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 67:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 132 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-818-111-67

Query Match 39.5%; Score 670; DB 4; Length 132;
 Best Local Similarity 99.2%; Pred. No. 4.6e-51;
 Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 199 TAASDNFQLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 258
 Db 1 TAASDNFQLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
 QY 259 VORVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVIVSVTWOTKSGGTR 318
 Db 61 VORVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVIVSVTWOTKSGGTR 120
 QY 319 TGNVTLAEGPPA 330
 Db 121 TGNVTLAEGPPA 132

RESULT 12
 US-09-056-556-66
 ; Sequence 66, Application US/09056556
 ; Patent No. 6350456
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
 ; NUMBER OF SEQUENCES: 241
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/056.556
 ; FILING DATE: 07-APR-1998
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 210121.457
 ; REFERENCE/DOCKET NUMBER: 210121.457
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 66:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 132 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-09-056-556-66

Query Match 39.5%; Score 670; DB 4; Length 132;
 Best Local Similarity 99.2%; Pred. No. 4.6e-51;
 Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 199 TAASDNFQLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 258
 Db 1 TAASDNFQLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
 QY 259 VORVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVIVSVTWOTKSGGTR 318
 Db 61 VORVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVIVSVTWOTKSGGTR 120
 QY 319 TGNVTLAEGPPA 330
 Db 121 TGNVTLAEGPPA 132

RESULT 13
 US-09-072-596-67
 ; Sequence 67, Application US/09072596
 ; Patent No. 6458366
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonia
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Twardzik, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; APPLICANT: Lodes, Michael J.

```

; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-596-67

Query Match 39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 4.6e-51;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 199 TAASDNFQLSQGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 258
Db 1 TAASDNFQLSQGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
QY 259 VQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 318
Db 61 VQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 120
QY 319 TGNVTLAEGPPA 330
Db 121 TGNVTLAEGPPA 132

RESULT 14
US-09-072-967-66
; Sequence 66, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Read, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington

```

```

; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-967-66

Query Match 39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 4.6e-51;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 199 TAASDNFQLSQGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 258
Db 1 TAASDNFQLSQGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
QY 259 VQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 318
Db 61 VQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISTVWTKSGGTR 120
QY 319 TGNVTLAEGPPA 330
Db 121 TGNVTLAEGPPA 132

RESULT 15
US-09-636-215-819
; Sequence 819, Application US/09636215
; Patent No. 6620922
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fauger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.42717C17
; CURRENT APPLICATION NUMBER: US/09/636,215
; CURRENT FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 852
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 819

```

```
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-636-215-819

Query Match      39.5%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 4.6e-51;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 199 TAASDNFQLSQGGGFAIPIGQAWAIAGQIRSGGGSTVHIGPTAFGLGLGVVDNNGNGAR 258
      |||
Db 1 TAASDNFQLSQGGGFAIPIGQAWAIAGQIRSGGGSTVHIGPTAFGLGLGVVDNNGNGAR 60

Qy 259 VQVVVGSAPAAASLGISTGDIVTAVDGPINSATAMADALNGHHPGDVIVTWOTKSGGTR 318
      |||
Db 61 VQVVVGSAPAAASLGISTGDIVTAVDGPINSATAMADALNGHHPGDVIVTWOTKSGGTR 120

Qy 319 TGNVTLAEGPPA 330
      |||
Db 121 TGNVTLAEGPPA 132

Search completed: June 22, 2004, 17:27:29
Job time : 12.78 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:44:25 ; Search time 32.2882 Seconds
(without alignments)
2301.458 Million cell updates/sec

Title: US-09-886-349A-12
Perfect score: 1306
Sequence: 1 VAWSVTAGQAEELTAQVRV.....YGHRRDGGKYAXSGRRNGGPA 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1304	99.8	263	AAW32379	Mycobacte
2	1304	99.8	263	AAW32447	Mycobacte
3	1304	99.8	263	AAW64317	Mycobacte
4	1304	99.8	263	AAW81680	M. tuberc
5	1304	99.8	263	AAW32062	Mycobacte
6	1304	99.8	263	AAW38984	M. tuberc
7	1304	99.8	263	AAW39121	M. tuberc
8	1304	99.8	263	AAE29706	Mycobacte
9	1304	99.8	263	AAE17570	Mycobacte
10	1304	99.8	358	AAU74591	Antigenic
11	1187	90.9	391	AAW32381	Mycobacte
12	1187	90.9	391	AAW32449	Mycobacte
13	1187	90.9	391	AAW64335	Mycobacte
14	1187	90.9	391	AAW81702	M. tuberc
15	1187	90.9	391	AAV04778	Mycobacte
16	1187	90.9	391	AAW38989	M. tuberc
17	1187	90.9	391	AAV39132	M. tuberc
18	1187	90.9	391	AAU01888	M. tuberc
19	1187	90.9	391	AAE29707	Mycobacte
20	1187	90.9	391	AAE17571	Mycobacte
21	1187	90.9	596	AAV32070	Mycobacte
22	1187	90.9	596	AAE29710	Mycobacte
23	1187	90.9	596	AAE17574	Mycobacte
24	1187	90.9	599	AAU74599	Antigenic
25	1187	90.9	600	AAV32068	Mycobacte

26	1187	90.9	600	5	AAU74597	Antigenic
27	1187	90.9	723	7	ADA26354	Mycobacte
28	1187	90.9	729	4	AAO22142	Ral2-H9-3
29	1187	90.9	729	5	AAE29709	Mycobacte
30	1187	90.9	729	5	AAE17573	Mycobacte
31	1187	90.9	729	7	ADA26374	Mycobacte
32	1187	90.9	744	4	AAU01902	M. tuberc
33	1187	90.9	788	4	AAU01903	M. tuberc
34	1187	90.9	813	7	ADA26367	Mycobacte
35	1187	90.9	815	4	AAU01904	M. tuberc
36	1187	90.9	825	7	ADA26366	Mycobacte
37	1187	90.9	875	7	ADA26365	Mycobacte
38	1187	90.9	930	5	AAE29731	Mycobacte
39	1187	90.9	930	7	ADA26364	Mycobacte
40	1187	90.9	1010	7	ADA26356	Mycobacte
41	1187	90.9	1016	7	ADA26370	M. bovis
42	1187	90.9	1022	7	ADA26369	Mycobacte
43	1187	90.9	1184	7	ADA26368	Mycobacte
44	1182	90.5	394	2	AAU04779	Mycobacte
45	1182	90.5	729	5	AAE29708	Mycobacte

ALIGNMENTS

RESULT 1
AAW32379
ID AAW32379 standard; protein; 263 AA.
XX
AC AAW32379;
XX
DT 13-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TbH-9.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT Misc-difference 254
FT /note= "Any amino acid"
XX
FN WO9709429-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US014675.
XX
PR 01-SEP-1995; 95US-00523435.
XX
PR 22-SEP-1995; 95US-00532136.
XX
PR 22-MAR-1996; 96US-00620280.
XX
PR 05-JUN-1996; 96US-00658800.
XX
PR 12-JUL-1996; 96US-00680573.
XX
(CORI-) CORIXA CORP.
XX
PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TH, Twardzik DR;
XX
DR WPI; 1997-192904/17.
XX
DR N-PSDB; AAT91432.
XX
FT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
FT useful for diagnosis of M. tuberculosis infection.
XX
PS Example 3; Page 138-139; 190pp; English.
XX
CC A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
CC variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a M.tuberculosis antigen,

CC TbH-9. The immunogenic polypeptide can be used to diagnose M.tuberculosis
 CC infection by forming complexes with specific antibodies in the sample.
 CC Fragments of DNA encoding the immunogenic polypeptide can be used as
 CC diagnostic primers or probes and agents that bind to the antigen,
 CC especially monoclonal antibodies or equivalent polyclonal antibodies, are
 CC also used for diagnosis

XX Sequence 263 AA;
 SQ

Query Match 99.8%; Score 1304; DB 2; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVIAENRAELMILITATNLLGQNTPA 60
 Db 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVIAENRAELMILITATNLLGQNTPA 60
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 120
 Db 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 120
 QY 121 TAAANQLMNNVPQALKQLAQPTQCTTPSSKLGGLWKTVPSPHRSPISNMVMANNHMSMTN 180
 Db 121 TAAANQLMNNVPQALKQLAQPTQCTTPSSKLGGLWKTVPSPHRSPISNMVMANNHMSMTN 180
 QY 181 SGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSSLGSSGLGGGVAANLGRAA 240
 Db 181 SGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSSLGSSGLGGGVAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 2

AAW32447
 ID AAW32447 standard; protein; 263 AA.

AC AAW32447;
 XX
 XX 09-JAN-1998 (first entry)
 DT
 XX
 DE Mycobacterium tuberculosis antigen TbH-9.
 XX
 XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 KW skin testing; M.tuberculosis.
 XX
 OS Mycobacterium tuberculosis.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 254
 FT /note= "Any amino acid"

XX WO9709428-A2.
 XX
 XX 13-MAR-1997.
 XX
 XX 30-AUG-1996; 96WO-US014674.
 XX
 XX 01-SEP-1995; 95US-00523436.
 PR 22-SEP-1995; 95US-00533634.
 PR 22-MAR-1996; 96US-00620874.
 PR 05-JUN-1996; 96US-00659683.
 PR 12-JUL-1996; 96US-00680574.

XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky Ya, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TH, Twardzik DR;
 XX WPI; 1997-192903/17.
 DR N-PSDB; AAT91496.
 XX

PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also for
 PT diagnosis.
 XX
 XX Example 3; Page 126-127; 168pp; English.
 XX
 CC A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
 CC variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis antigen,
 CC TbH-9. The immunogenic protein, and fusion proteins containing one or
 CC more of the proteins or one of the proteins plus ESAT-6, are useful in
 CC vaccines, preferably when formulated with a non-specific adjuvant, to
 CC induce an immune response against M.tuberculosis (for treatment or
 CC prevention)
 XX
 SQ Sequence 263 AA;

Query Match 99.8%; Score 1304; DB 2; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVIAENRAELMILITATNLLGQNTPA 60
 Db 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPVIAENRAELMILITATNLLGQNTPA 60
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 120
 Db 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 120
 QY 121 TAAANQLMNNVPQALKQLAQPTQCTTPSSKLGGLWKTVPSPHRSPISNMVMANNHMSMTN 180
 Db 121 TAAANQLMNNVPQALKQLAQPTQCTTPSSKLGGLWKTVPSPHRSPISNMVMANNHMSMTN 180
 QY 181 SGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSSLGSSGLGGGVAANLGRAA 240
 Db 181 SGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSSLGSSGLGGGVAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 3

AAW64317
 ID AAW64317 standard; protein; 263 AA.

AC AAW64317;
 XX
 XX 17-OCT-2003 (revised)
 DT 09-NOV-1998 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen TbH-9.
 XX
 XX Tuberculosis; infection; diagnosis; antigen; TbH-9.
 OS Mycobacterium tuberculosis; strain H37Rv.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 254
 FT /note= "encoded by NAG"
 XX WO9816645-A2.
 XX
 XX 23-APR-1998.
 XX
 XX 07-OCT-1997; 97WO-US018214.
 XX
 XX 11-OCT-1996; 96US-00729622.
 PR 13-MAR-1997; 97US-00618111.
 XX (CORI-) CORIXA CORP.
 XX

PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX
 DR WPI: 1998-251292/22.
 DR N-PSDB; AAV44371.
 XX

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and
 PT diagnosis of tuberculosis.
 PT

XX Example 3; Page 125-126; 250pp; English.
 PS

XX This polypeptide comprises Mycobacterium tuberculosis antigen TbH-9. A
 CC DNA sequence (see AAV44371) coding for antigen TbH-9 was isolated from a
 CC M. tuberculosis strain H37Rv expression library using sera from patients
 CC having pulmonary or pleural tuberculosis. The invention relates to
 CC compositions and methods for diagnosing tuberculosis. It provides
 CC polypeptides (see AAV64291-W64379) comprising an antigenic portion of a
 CC soluble M. tuberculosis antigen, or an immunogenic portion of an M.
 CC tuberculosis antigen, as well as DNA sequences encoding such
 CC polypeptides, recombinant expression vectors and transformed or
 CC transfected host cells. Also claimed are methods and diagnostic kits for
 CC detecting M. tuberculosis infection in a patient using these
 CC polypeptides, antibodies or oligonucleotide probes and primers, for the
 CC diagnosis of tuberculosis. (Updated on 17-OCT-2003 to standardise OS
 CC field)
 XX

XX Sequence 263 AA;
 SQ

Query Match 99.8%; Score 1304; DB 2; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAWMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
 DB 1 VAWMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATALLPPEEAPEMTSAGGLEQAAVEEASD 120
 DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATALLPPEEAPEMTSAGGLEQAAVEEASD 120
 QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMAHHMSMTN 180
 DB 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMAHHMSMTN 180
 QY 181 SGVSWTNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSLSGLSGGLGGVAAANLGRAA 240
 DB 181 SGVSWTNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSLSGLSGGLGGVAAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 4
 AAW81680
 ID AAW81680 standard; protein; 263 AA.
 XX
 AC AAW81680;
 XX

DT 27-JAN-1999 (first entry)
 XX

XX M. tuberculosis immunogenic polypeptide TbH-9.
 XX

XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 KW vaccine; pharmaceutical; infection; diagnosis.
 KW

XX Mycobacterium tuberculosis.
 OS

XX Key Location/Qualifiers
 XX Misc-difference 254
 FT /label= unknown
 FT
 XX

BN WO9816646-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US018293.
 XX
 PR 11-OCT-1996; 96US-00730510.
 PR 13-MAR-1997; 97US-00818112.
 XX
 PA (CORI-) CORIXA CORP.
 XX

XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX

XX WPI: 1998-261042/23.
 DR N-PSDB; AAV64479.
 DR

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and for
 PT diagnosis, treatment and prevention of tuberculosis.
 PT

XX Example 3b; Page 119-120; 230pp; English.
 PS

XX This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
 CC inducing protective immunity against tuberculosis (TB). This sequence can
 CC be formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis
 XX

XX Sequence 263 AA;
 SQ

Query Match 99.8%; Score 1304; DB 2; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAWMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
 DB 1 VAWMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEEAPEMTSAGGLEQAAVEEASD 120
 DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEEAPEMTSAGGLEQAAVEEASD 120
 QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMAHHMSMTN 180
 DB 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMAHHMSMTN 180
 QY 181 SGVSWTNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSLSGLSGGLGGVAAANLGRAA 240
 DB 181 SGVSWTNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSLSGLSGGLGGVAAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 5
 AAY32062
 ID AAY32062 standard; protein; 263 AA.
 XX
 AC AAY32062;
 XX

DT 17-JAN-2000 (first entry)
 XX

XX Mycobacterium tuberculosis antigen TbH9.
 XX

XX Tuberculosis; antigen; fusion protein; TbH9; diagnosis; therapy; vaccine;
 KW immunogen.
 KW

XX Mycobacterium tuberculosis.
 OS

XX Key Location/Qualifiers
 FH

FT Misc-difference 254
 XX /note= "not identified"
 XX WO9951748-A2.
 XX 14-OCT-1999.
 XX 07-APR-1999; 99WO-US007717.
 XX 07-APR-1998; 98US-00056556.
 XX 30-DEC-1998; 98US-00223040.
 XX (CORI-) CORIXA CORP.
 XX
 XX Skeiky YAW, Alderson M, Campos-Neto A;
 XX WPI; 1999-601610/51.
 XX
 XX New fusion proteins useful for diagnosis, prevention and treatment of
 XX tuberculosis.
 XX
 XX Claim 1; Fig 4C-D; 83pp; English.
 XX
 XX This sequence represents the Mycobacterium tuberculosis antigen TbH9. The
 XX invention provides fusion proteins (see AAY32059-71) containing at least
 XX 2 M. tuberculosis antigens such as TbH9, e.g. Mtb32A (see AAY32059) and a
 XX TbH9-Tb38-1 fusion. The new fusion proteins are useful as vaccines for
 XX preventing tuberculosis (claimed), for diagnosis (via in vitro assays or
 XX intradermal skin tests for detection of anti-M. tuberculosis antibodies),
 XX monitoring of disease progression, and treatment of tuberculosis. They
 XX are more effective immunogens than mixtures of the individual protein
 XX components
 XX
 XX Sequence 263 AA;
 SQ
 Query Match 99.8%; Score 1304; DB 2; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAMSVTAGQAEILTAACQVRVAAAYAYTAYGLTVPPVPIAENRAELMILITATNLLGQNTPA 60
 Db 1 VAMSVTAGQAEILTAACQVRVAAAYAYTAYGLTVPPVPIAENRAELMILITATNLLGQNTPA 60
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPEMTSAGGLLEQAAVEEASD 120
 Db 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPEMTSAGGLLEQAAVEEASD 120
 QY 121 TAAANQLMNNVPQALKOLAQPTGTTTSSKLGGLWKTVSPHRSPISNMVSMANNHSMNTN 180
 Db 121 TAAANQLMNNVPQALKOLAQPTGTTTSSKLGGLWKTVSPHRSPISNMVSMANNHSMNTN 180
 QY 181 SGVSMNTNTLSSMLKGFAPAAAQAQVTAACQNGVRAMSSLGSSGLGGVAAANLGRAA 240
 Db 181 SGVSMNTNTLSSMLKGFAPAAAQAQVTAACQNGVRAMSSLGSSGLGGVAAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 6
 AAY38984
 ID AAY38984 standard; protein; 263 AA.
 AC AAY38984;
 XX
 XX 05-NOV-1999 (first entry)
 DT
 XX M. tuberculosis recombinant antigen protein TbH-9.
 XX Antigen; diagnosis; detection; infection; antibody; immunisation;
 XX vaccine; immunity.
 XX

OS Mycobacterium tuberculosis.
 XX WO9942118-A2.
 XX 26-AUG-1999.
 XX 17-FEB-1999; 99WO-US003265.
 XX 18-FEB-1998; 98US-00024753.
 XX 05-MAY-1998; 98US-00072596.
 XX (CORI-) CORIXA CORP.
 XX
 XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 XX Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
 XX WPI; 1999-527416/44.
 XX N-PSDB; AAZ19069.
 XX
 XX New polypeptide comprising antigenic portions of M. tuberculosis.
 XX
 XX Example 3; Page 160-161; 323pp; English.
 XX
 XX This invention describes novel recombinant antigens and their encoding
 XX nucleic acids derived from Mycobacterium tuberculosis. The novel
 XX polypeptides are useful for detecting M. tuberculosis infection in a
 XX biological sample by detecting antibodies which bind with the
 XX polypeptides, and are useful as vaccines for immunizing against M.
 XX tuberculosis infection. The new detection methods are needed as current
 XX vaccination strategies do not provide 100% immunity
 XX
 XX Sequence 263 AA;
 SQ
 Query Match 99.8%; Score 1304; DB 2; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAMSVTAGQAEILTAACQVRVAAAYAYTAYGLTVPPVPIAENRAELMILITATNLLGQNTPA 60
 Db 1 VAMSVTAGQAEILTAACQVRVAAAYAYTAYGLTVPPVPIAENRAELMILITATNLLGQNTPA 60
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPEMTSAGGLLEQAAVEEASD 120
 Db 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPEMTSAGGLLEQAAVEEASD 120
 QY 121 TAAANQLMNNVPQALKOLAQPTGTTTSSKLGGLWKTVSPHRSPISNMVSMANNHSMNTN 180
 Db 121 TAAANQLMNNVPQALKOLAQPTGTTTSSKLGGLWKTVSPHRSPISNMVSMANNHSMNTN 180
 QY 181 SGVSMNTNTLSSMLKGFAPAAAQAQVTAACQNGVRAMSSLGSSGLGGVAAANLGRAA 240
 Db 181 SGVSMNTNTLSSMLKGFAPAAAQAQVTAACQNGVRAMSSLGSSGLGGVAAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 RESULT 7
 AAY39121
 ID AAY39121 standard; protein; 263 AA.
 AC AAY39121;
 XX
 XX 05-NOV-1999 (first entry)
 DT
 XX M. tuberculosis antigen TbH-9 amino acid sequence.
 XX
 XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 XX immunotherapy; diagnosis; immunisation; vaccine; infection;
 XX immune response; skin test.
 XX
 XX Mycobacterium tuberculosis.
 OS


```

XX PN WO9942076-A2.
XX PD 26-AUG-1999.
XX PF 17-FEB-1999; 99WO-US003268.
XX PR 18-FEB-1998; 98US-00025197.
XX PR 05-MAY-1998; 98US-00072967.
XX PA (CORI-) CORIXA CORP.
XX PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX DR WPI; 1999-527409/44.
XX PT New antigens from Mycobacterium tuberculosis useful in diagnostic skin
XX PT tests and protective or therapeutic vaccines or compositions.
XX PS Example 3; Page 115; 299pp; English.
XX CC The present invention describes polypeptides comprising an immunogenic
XX CC part of a Mycobacterium tuberculosis antigen (Ag). Also described are
XX CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
XX CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
XX CC polypeptides fragments, can be used in pharmaceutical compositions or
XX CC vaccines to generate a protective or therapeutic immune response to M.
XX CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
XX CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
XX CC killer cells and/or macrophages in tuberculosis-immune subjects. AA219249
XX CC to AA219460 and AA39083 to AA39225 are used in the exemplification of
XX CC the present invention
XX SQ Sequence 263 AA;

Query Match          99.8%; Score 1304; DB 2; Length 263;
Best Local Similarity 100.0%; Pred. No. 3.1e-105; Indels 0; Gaps 0;
Matches 263; Conservative 0; Mismatches 0;

QY 1 VAMSVTAGQAEATAAQVRVAAAAYETAYGLTVPVPIAENRAELMILITNLLGQNTPA 60
DB 1 VAMSVTAGQAEATAAQVRVAAAAYETAYGLTVPVPIAENRAELMILITNLLGQNTPA 60

QY 61 IAYNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120
DB 61 IAYNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120

QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN 180
DB 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN 180

QY 181 SGVSMNTLTSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240
DB 181 SGVSMNTLTSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240

QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 8
AAE29706
ID AAE29706 standard; protein; 263 AA.
XX AC AAE29706;
XX DT 27-JAN-2003 (first entry)
XX DE Mycobacterium sp. TbH9 antigenic protein.
XX KW Vaccine; immunity; diagnostic agent; gene therapy; TbH9 antigen.
XX
```

```

OS Mycobacterium sp.
XX Key Location/Qualifiers
XX FT Misc-difference 254
XX FT /note= "Encoded by NAG"
XX PN WO200272792-A2.
XX PD 19-SEP-2002.
XX PF 13-MAR-2002; 2002WO-US008223.
XX PR 13-MAR-2001; 2001US-0275837P.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky Y, Brannon M, Guderian J;
XX DR WPI; 2002-759844/82.
XX DR N-PSDB; AAD47081.
XX PT New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif,
XX PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
XX PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
XX PT tuberculosis.
XX PS Disclosure; Page 84; 155pp; English.
XX CC The invention relates to a recombinant nucleic acid molecule encoding a
XX CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
XX CC polynucleotide sequence encoding an antigen or an antigenic fragment from
XX CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
XX CC polypeptide or its fragment. The Leishmania polynucleotide is selected
XX CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention
XX CC are used in methods for eliciting immune response in mammals. They are
XX CC useful as vaccines to elicit protective immunity against pathogenic
XX CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
XX CC polypeptides are used for enhancing the expression of polynucleotides, as
XX CC in vivo diagnostic agents and for raising antibodies in a non-human
XX CC animal. The invention is used in gene therapy. The present sequence is
XX CC Mycobacterium sp. TBH9 antigenic protein
XX SQ Sequence 263 AA;

Query Match          99.8%; Score 1304; DB 5; Length 263;
Best Local Similarity 100.0%; Pred. No. 3.1e-105; Indels 0; Gaps 0;
Matches 263; Conservative 0; Mismatches 0;

QY 1 VAMSVTAGQAEATAAQVRVAAAAYETAYGLTVPVPIAENRAELMILITNLLGQNTPA 60
DB 1 VAMSVTAGQAEATAAQVRVAAAAYETAYGLTVPVPIAENRAELMILITNLLGQNTPA 60

QY 61 IAYNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120
DB 61 IAYNEAEGEMWAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120

QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN 180
DB 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSMANNHMSMTN 180

QY 181 SGVSMNTLTSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240
DB 181 SGVSMNTLTSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240

QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 9
AAE17570
ID AAE17570 standard; protein; 263 AA.
XX
```

AC AAE17570;
 XX 22-APR-2002 (first entry)
 XX Mycobacterium species MTB39 (TbH9) protein #1.
 DE Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB39; TbH9 protein.
 XX Mycobacterium sp.
 OS
 XX Key Location/Qualifiers
 XX Misc-difference 254 /label= Unknown
 FT /note= "Encoded by NAG"
 FT
 XX WO200198460-A2.
 XX 27-DEC-2001.
 XX 20-JUN-2001; 2001WO-US019959.
 PF 20-JUN-2000; 2000US-00597796.
 PR 01-FEB-2001; 2001US-0265737P.
 XX (CORI-) CORIXA CORP.
 XX Skeiky Y, Reed S, Alderson M;
 PI WPI; 2002-147798/19.
 XX N-PSDB; AAD28340.
 XX Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 XX Claim 83; Page 100; 136pp; English.
 XX The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB39 (TbH9) protein
 XX Sequence 263 AA:
 Query Match 99.8%; Score 1304; DB 5; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105; Gaps 0;
 Matches 263; Conservative 0; Mismatches 0; Indels 0;
 QY 1 VAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPPPVIAENRAELMILIATNLGNTPA 60
 Db 1 VAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPPPVIAENRAELMILIATNLGNTPA 60
 QY 61 IAVNEAYGEMWAQDAAMFGYAAATATATATLLPPEEAPMTSAGLLEQAAVEEASD 120
 Db 61 IAVNEAYGEMWAQDAAMFGYAAATATATATLLPPEEAPMTSAGLLEQAAVEEASD 120

QY 121 TAAANQLMNNVPQALKQLAOPQTGTPSSKLGGLWKTVPSPHSPISNMVSNNHMSMTN 180
 Db 121 TAAANQLMNNVPQALKQLAOPQTGTPSSKLGGLWKTVPSPHSPISNMVSNNHMSMTN 180
 QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLSGLSSGSLGGGVAANLGRAA 240
 Db 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLSGLSSGSLGGGVAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 RESULT 10
 AAU74591
 ID AAU74591 standard; protein; 358 AA.
 XX AAU74591;
 XX 29-AUG-2003 (revised)
 DT 08-MAY-2002 (first entry)
 XX Antigenic fusion protein TbH9-Tb38-1.
 DE Fusion protein; tuberculosis; Mycobacterium tuberculosis;
 KW tuberculostatic; immunogen; vaccine; TbH9-Tb38-1; TbH9; Tb38-1.
 XX Mycobacterium tuberculosis.
 OS Chimeric.
 XX Key Location/Qualifiers
 FT Misc-difference 254 /label= unknown
 XX US2002009459-A1.
 XX 24-JAN-2002.
 XX 07-APR-1999; 99US-00287849.
 PR 13-MAR-1997; 97US-00818112.
 PR 01-OCT-1997; 97US-00942578.
 PR 18-FEB-1998; 98US-00025197.
 PR 07-APR-1998; 98US-00056556.
 PR 30-DEC-1998; 98US-00223040.
 XX (REED/) REED S G.
 PA (SKEI/) SKEIKY Y A.
 PA (DILL/) DILLON D C.
 PA (ALDE/) ALDERSON M.
 PA (CAMPE/) CAMPOS-NETO A.
 XX Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
 WPI; 2002-171134/22.
 XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for
 PT diagnosing, treating or preventing M. tuberculosis infection,
 PT particularly as vaccine for treating or preventing tuberculosis.
 XX Claim 1; Fig 4C-D; 62pp; English.
 XX The invention relates to a purified polypeptide which induces an immune
 CC response of Mycobacterium tuberculosis. Polypeptides of the invention are
 CC useful for diagnosing, treating or preventing M. tuberculosis infection,
 CC particularly tuberculosis infection. In particular, the polypeptides are
 CC useful as a vaccine formulation with an adjuvant to afford long-term
 CC protection in animals against the development of tuberculosis. The
 CC protein coding sequence may be used to encode a protein product for use
 CC as an immunogen to induce and/or enhance an immune response to M.
 CC tuberculosis. This sequence represents an M. tuberculosis fusion protein
 CC of the invention. Note: The specification states that this polypeptide is
 CC encoded by the polynucleotide shown in ABK14131. (Updated on 29-AUG-2003

CC to standardise OS field)

XX Sequence 358 AA;

SQ Query Match 99.8%; Score 1304; DB 5; Length 358;
Best Local Similarity 100.0%; Pred. No. 4.7e-105; Indels 0; Gaps 0;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWSVTAGQAEHTAAQVRVAAAYETAYGLTVPPVIAENRAELMILIAITNLLGQNTPA 60
DB 1 VAWSVTAGQAEHTAAQVRVAAAYETAYGLTVPPVIAENRAELMILIAITNLLGQNTPA 60

QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 120
DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 120

QY 121 TAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHMTN 180
DB 121 TAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHMTN 180

QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSGLGGGVAANLGRAA 240
DB 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSGLGGGVAANLGRAA 240

QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 11
AAW32381
ID AAW32381 standard; protein; 391 AA.
XX AC AAW32381;
XX DT 13-JAN-1998 (first entry)
XX DE Mycobacterium tuberculosis antigen TbH-9FL.
XX KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
XX SKW skin testing; M.tuberculosis.
XX OS Mycobacterium tuberculosis.
XX PN WO9709429-A2.
XX PD 13-MAR-1997.
XX PF 30-AUG-1996; 96WO-US014675.
XX PR 01-SEP-1995; 95US-00523435.
XX PR 22-SEP-1995; 95US-00532136.
XX PR 22-MAR-1996; 96US-00620280.
XX PR 05-JUN-1996; 96US-00658800.
XX PR 12-JUL-1996; 96US-00680573.
XX PA (CORI-) CORIXA CORP.
XX PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX PI Vedvick TH, Twardzik DR;
XX DR WPI; 1997-192904/17.
XX DR N-PSDB; AAT91455.
XX PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
XX PT useful for diagnosis of M. tuberculosis infection.
XX PS Example 3; Page 150-152; 190pp; English.
XX CC A new immunogenic polypeptide has been developed comprising an
XX CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
XX CC variant differing only in conservative substitutions and/or
XX CC modifications). The present sequence represents a M.tuberculosis antigen,

CC TbH-9FL The immunogenic polypeptide can be used to diagnose
CC M.tuberculosis infection by forming complexes with specific antibodies in
CC the sample. Fragments of DNA encoding the immunogenic polypeptide can be
CC used as diagnostic primers or probes and agents that bind to the antigen,
CC especially monoclonal antibodies or equivalent polyclonal antibodies, are
CC also used for diagnosis

XX SQ Sequence 391 AA;
Query Match 90.9%; Score 1187; DB 2; Length 391;
Best Local Similarity 99.6%; Pred. No. 8.2e-95;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWSVTAGQAEHTAAQVRVAAAYETAYGLTVPPVIAENRAELMILIAITNLLGQNTPA 60
DB 74 VAWSVTAGQAEHTAAQVRVAAAYETAYGLTVPPVIAENRAELMILIAITNLLGQNTPA 133

QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 120
DB 134 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 193

QY 121 TAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHMTN 180
DB 194 TAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHMTN 253

QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSGLGGGVAANLGRAA 240
DB 254 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSGLGGGVAANLGRAA 313

QY 241 SV 242
DB 314 SV 315

RESULT 12
AAW32449
ID AAW32449 standard; protein; 391 AA.
XX AC AAW32449;
XX DT 09-JAN-1998 (first entry)
XX DE Mycobacterium tuberculosis antigen TbH-9FL.
XX KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
XX SKW skin testing; M.tuberculosis.
XX OS Mycobacterium tuberculosis.
XX PN WO9709428-A2.
XX PD 13-MAR-1997.
XX PF 30-AUG-1996; 96WO-US014674.
XX PR 01-SEP-1995; 95US-00523436.
XX PR 22-SEP-1995; 95US-00533634.
XX PR 22-MAR-1996; 96US-00620874.
XX PR 05-JUN-1996; 96US-00659683.
XX PR 12-JUL-1996; 96US-00680574.
XX PA (CORI-) CORIXA CORP.
XX PI Reed SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;
XX PI Vedvick TH, Twardzik DR;
XX DR WPI; 1997-192903/17.
XX DR N-PSDB; AAT91521.
XX PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
XX PT useful in vaccines for prevention or treatment of tuberculosis, also for
XX PT diagnosis.


```

XX WPI: 1998-261042/23.
DR N-PSDB; AAV64503.
XX
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
PT develop products for the detection of M. tuberculosis infection and for
PT diagnosis, treatment and prevention of tuberculosis.
XX
XX Example 3B; Page 128-129; 230pp; English.
XX
XX This sequence represents an immunogenic portion of a soluble
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
CC inducing protective immunity against tuberculosis (TB). This sequence can
CC be formulated into vaccines and/or pharmaceutical compositions for
CC immunising against M. tuberculosis infection or may be used for the
CC diagnosis of tuberculosis
XX
XX Sequence 391 AA;
SQ

```

Query Match 90.9%; Score 1187; DB 2; Length 391;
 Best Local Similarity 99.6%; Pred. No. 8.2e-95;
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPFPVIAENRAELMILLIATNLLGQNTPA 60
DB 74 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPFPVIAENRAELMILLIATNLLGQNTPA 133
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 120
DB 134 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 193
QY 121 TAAANQLMNNVPQALKQLAQPTGTTSPSSKLGGLWKTVPSPHSPISNMYSMANNHSMTN 180
DB 194 TAAANQLMNNVPQALKQLAQPTGTTSPSSKLGGLWKTVPSPHSPISNMYSMANNHSMTN 253
QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAAGNGVRAMSSLGSSLGSSGGLGGVAAANLGRAA 240
DB 254 SGVSMNTLSSMLKGFAPAAAQAVQTAAGNGVRAMSSLGSSLGSSGGLGGVAAANLGRAA 313
QY 241 SV 242
DB 314 SV 315

```

RESULT 15

```

AY04778
ID AAY04778 standard; protein; 391 AA.
XX
AC AAY04778;
XX

```

```

06-JUL-1999 (first entry)
XX

```

```

DE Mycobacterium species protein sequence 5R.
XX

```

```

XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW hybridisation; detection; vaccine; immunisation; infection.
XX

```

```

OS Mycobacterium sp.
XX

```

```

XX WO9909186-A2.
PN

```

```

PD 25-FEB-1999.
XX

```

```

XX 14-AUG-1998; 98WO-FR001813.
XX

```

```

PR 14-AUG-1997; 97FR-00010404.
XX

```

```

PR 11-SEP-1997; 97FR-00011325.
XX

```

```

XX (INSP ) INST PASTEUR.
PA

```

```

XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
PI

```

```

PI Goguet De La Salmoniere Y;
XX

```

```

DR WPI: 1999-181045/15.
DR N-PSDB; AAX34030.
XX

```

```

PT Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in infection-associated
PT protein expression.
XX

```

```

PS Claim 32; Fig 5R; 309pp; French.
XX

```

```

CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
CC from various Mycobacterium species microorganisms. The encoding
CC nucleotide sequences can be used as primers and probes for methods for
CC detecting and identifying mycobacteria, especially belonging to the M.
CC tuberculosis complex. The encoded proteins can be used in vaccines for
CC immunisation against a bacterial or viral infection
XX

```

```

XX Sequence 391 AA;
SQ

```

```

Query Match 90.9%; Score 1187; DB 2; Length 391;
Best Local Similarity 99.6%; Pred. No. 8.2e-95;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPFPVIAENRAELMILLIATNLLGQNTPA 60
DB 74 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPFPVIAENRAELMILLIATNLLGQNTPA 133
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 120
DB 134 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAAVEEASD 193
QY 121 TAAANQLMNNVPQALKQLAQPTGTTSPSSKLGGLWKTVPSPHSPISNMYSMANNHSMTN 180
DB 194 TAAANQLMNNVPQALKQLAQPTGTTSPSSKLGGLWKTVPSPHSPISNMYSMANNHSMTN 253
QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAAGNGVRAMSSLGSSLGSSGGLGGVAAANLGRAA 240
DB 254 SGVSMNTLSSMLKGFAPAAAQAVQTAAGNGVRAMSSLGSSLGSSGGLGGVAAANLGRAA 313
QY 241 SV 242
DB 314 SV 315

```

Search completed: June 22, 2004, 17:15:28

Job time : 33.2882 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:07:54 ; Search time 9.38833 Seconds
(without alignments)
1446.225 Million cell updates/sec

Title: US-09-886-349A-12
Perfect score: 1306
Sequence: 1 VAMSVTAQAEILTAQVRV.....YGRDGGKYXSGRRNGGPA 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/6CTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1304	99.8	263	3	US-08-818-112-91
2	1304	99.8	263	4	US-08-818-111-92
3	1304	99.8	263	4	US-09-056-556-91
4	1304	99.8	263	4	US-09-072-596-92
5	1304	99.8	263	4	US-09-072-967-91
6	1304	99.8	358	4	US-09-287-849-8
7	1187	90.9	391	3	US-08-818-112-107
8	1187	90.9	391	4	US-08-818-111-102
9	1187	90.9	391	4	US-09-056-556-107
10	1187	90.9	391	4	US-09-072-596-102
11	1187	90.9	391	4	US-09-072-967-107
12	1187	90.9	596	4	US-09-287-849-26
13	1187	90.9	600	4	US-09-287-849-22
14	1182	90.5	729	4	US-09-223-040-2
15	1182	90.5	729	4	US-09-287-849-2
16	1001.5	76.7	396	3	US-08-818-112-111
17	1001.5	76.7	396	4	US-08-818-111-106
18	1001.5	76.7	396	4	US-09-056-556-111
19	1001.5	76.7	396	4	US-09-072-596-106
20	1001.5	76.7	396	4	US-09-072-967-111
21	949.5	72.7	359	3	US-08-818-111-109
22	949.5	72.7	359	4	US-08-818-111-104
23	949.5	72.7	359	4	US-09-056-556-109
24	949.5	72.7	359	4	US-09-072-596-104
25	949.5	72.7	359	4	US-09-072-967-109
26	429.5	32.9	400	4	US-09-073-009-126
27	429.5	32.9	400	4	US-09-073-010-126

28	325.5	24.9	423	4	US-09-073-009-142	Sequence 142, App
29	325.5	24.9	423	4	US-09-073-010-142	Sequence 142, App
30	325.5	24.9	710	4	US-09-287-849-16	Sequence 16, Appl
31	325.5	24.9	856	4	US-09-287-849-12	Sequence 12, Appl
32	320.5	24.5	943	4	US-09-477-135A-131	Sequence 131, Appl
33	239	18.3	204	4	US-08-311-731A-57	Sequence 57, Appl
34	235	18.0	208	4	US-08-311-731A-208	Sequence 208, App
35	223	17.1	141	4	US-09-073-009-15	Sequence 15, Appl
36	223	17.1	141	4	US-09-073-010-15	Sequence 15, Appl
37	197.5	15.1	943	4	US-09-056-556-204	Sequence 204, App
38	197.5	15.1	943	4	US-09-072-596-199	Sequence 199, App
39	197.5	15.1	943	4	US-09-072-967-204	Sequence 204, App
40	188.5	14.4	368	3	US-08-818-112-114	Sequence 114, App
41	188.5	14.4	368	4	US-08-818-111-109	Sequence 109, App
42	188.5	14.4	368	4	US-09-056-556-114	Sequence 114, App
43	188.5	14.4	368	4	US-09-072-596-109	Sequence 109, App
44	188.5	14.4	368	4	US-09-072-967-114	Sequence 114, App
45	188.5	14.4	371	4	US-09-050-739-92	Sequence 92, Appl

ALIGNMENTS

RESULT 1
US-08-818-112-91
; Sequence 91, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-91

Query Match 99.8%; Score 1304; DB 3; Length 263;
Best Local Similarity 100.0%; Pred. No. 4.8e-116;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 121 TAAANQLMNNVPOALKQLAQPTGTTTPSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180
Qy 181 SGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSSLGSSGLGGVAAANLGRAA 240
Db 181 SGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSSLGSSGLGGVAAANLGRAA 240
Qy 241 SVRYGHRDGGKYAXSGRRNGGPA 263
Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 4
US-09-072-596-92
; Sequence 92, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-596-92

Query Match 99.8%; Score 1304; DB 4; Length 263;
Best Local Similarity 100.0%; Pred. No. 4.8e-116;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VAWSVTAGAELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMLLIATNLLGNTPA 60
Db 1 VAWSVTAGAELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMLLIATNLLGNTPA 60
Qy 61 IAVNEAYGEMWAQDAAMFGYAAATATATATATLLPFEAPEMTSAGLLLEQAAAVEASD 120
Db 61 IAVNEAYGEMWAQDAAMFGYAAATATATATATLLPFEAPEMTSAGLLLEQAAAVEASD 120
Qy 121 TAAANQLMNNVPOALKQLAQPTGTTTPSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180
Db 121 TAAANQLMNNVPOALKQLAQPTGTTTPSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180

Qy 181 SGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSSLGSSGLGGVAAANLGRAA 240
Db 181 SGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSSLGSSGLGGVAAANLGRAA 240
Qy 241 SVRYGHRDGGKYAXSGRRNGGPA 263
Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 5
US-09-072-967-91
; Sequence 91, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-967-91

Query Match 99.8%; Score 1304; DB 4; Length 263;
Best Local Similarity 100.0%; Pred. No. 4.8e-116;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VAWSVTAGAELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMLLIATNLLGNTPA 60
Db 1 VAWSVTAGAELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMLLIATNLLGNTPA 60
Qy 61 IAVNEAYGEMWAQDAAMFGYAAATATATATATLLPFEAPEMTSAGLLLEQAAAVEASD 120
Db 61 IAVNEAYGEMWAQDAAMFGYAAATATATATATLLPFEAPEMTSAGLLLEQAAAVEASD 120
Qy 121 TAAANQLMNNVPOALKQLAQPTGTTTPSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180
Db 121 TAAANQLMNNVPOALKQLAQPTGTTTPSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180

QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSGLSSGLGGVAAANLGRAA 240
DB 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSGLSSGLGGVAAANLGRAA 240
QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 6

US-09-287-849-8
; Sequence 8, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-00920US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TB9-Tb38-1
; NAME/KEY: MOD RES
; LOCATION: (254)
; OTHER INFORMATION: Xaa = any amino acid
US-09-287-849-8

Query Match 99.8%; Score 1304; DB 4; Length 358;
Best Local Similarity 100.0%; Pred. No. 7.4e-116;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
DB 1 VAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATATLLPFEAPEMTSAGLLEQAAAVEASD 120
DB 61 IAVNEAEYGENWAQDAAMFGYAAATATATATATLLPFEAPEMTSAGLLEQAAAVEASD 120
QY 121 TAAANQLMNNVPQALKQLAQFTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 180
DB 121 TAAANQLMNNVPQALKQLAQFTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 180
QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSGLSSGLGGVAAANLGRAA 240
DB 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSGLSSGLGGVAAANLGRAA 240
QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 7

US-08-818-112-107
; Sequence 107, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-107

Query Match 90.9%; Score 1187; DB 3; Length 391;
Best Local Similarity 99.6%; Pred. No. 1.1e-104;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
DB 74 VAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 133
QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLEQAAAVEASD 120
DB 134 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLEQAAAVEASD 193
QY 121 TAAANQLMNNVPQALKQLAQFTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 180
DB 194 TAAANQLMNNVPQALKQLAQFTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 253
QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSGLSSGLGGVAAANLGRAA 240
DB 254 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSGLSSGLGGVAAANLGRAA 313
QY 241 SV 242
DB 314 SV 315

RESULT 8

US-08-818-111-102

; Sequence 102, Application US/08818111

; Patent No. 6338852

; GENERAL INFORMATION:

; APPLICANT: Read, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Campos-Neto, Antonia

; APPLICANT: Houghton, Raymond

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Twardzik, Daniel R.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS

; NUMBER OF SEQUENCES: 148

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/818,111

; FILING DATE: 13-MAR-1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.417C6

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 102:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 391 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-818-111-102

Query Match 90.9%; Score 1187; DB 4; Length 391;

Best Local Similarity 99.6%; Pred. No. 1.1e-104;

Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VAMSVTAGQAEITAAQVRVAAAYTAYGLTVPPFVIAENRAELMILIATNLLGONTPA 60

Db 74 VAMSVTAGQAEITAAQVRVAAAYTAYGLTVPPFVIAENRAELMILIATNLLGONTPA 133

Qy 61 IAVNEAEYGEWQAQDAAMFGYAAATATATATATLLPFEAPEMTSAAGLLEQAAAVEEASD 120

Db 134 IAVNEAEYGEWQAQDAAMFGYAAATATATATLLPFEAPEMTSAAGLLEQAAAVEEASD 193

Qy 121 TAAANQLMNVNPQALKOLAQPTQCTTPSSKLGGLWKTVPSPHSPISNVSMANNHSMNTN 180

Db 194 TAAANQLMNVNPQALQOLAQPTQCTTPSSKLGGLWKTVPSPHSPISNVSMANNHSMNTN 253

Qy 181 SGVSMNTNLTLSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSIGSSGLGGVAAANLGRAA 240

Db 254 SGVSMNTNLTLSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSIGSSGLGGVAAANLGRAA 313

Qy 241 SV 242

Db 314 SV 315

RESULT 9

US-09-056-556-107

; Sequence 107, Application US/09056556

TREATM

Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056.556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-056-556-107

Query Match 90.9%; Score 1187; DB 4; Length 391;
Best Local Similarity 99.6%; Pred. No. 1.le-104;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	VAMSVVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMLIATNLLGNTPA	60
Dd	74	VAMSVVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMLIATNLLGNTPA	133
Qy	61	IAVNEAEEYGENWAQDAAMFGYAAATATATATATLPPFEAPEMTSAGGLEEQAAAIVEASD	120
Dd	134	IAVNEAEEYGENWAQDAAMFGYAAATATATATATLPPFEAPEMTSAGGLEEQAAAIVEASD	193
Qy	121	TAAANQLMNVPQALKQAQPTGGTTPSKLGGLWKTVPSPHRSPISNMYSMANHHSMTN	180
Dd	194	TAAANQLMNVPQALKQAQPTGGTTPSKLGGLWKTVPSPHRSPISNMYSMANHHSMTN	253
Qy	181	SGVSWMTNTLSNMLKGFAPAAAAQAVQTAAQNGVRAMSSILGSSSLGGSGVGVAANLGRAA	240
Dd	254	SGVSWMTNTLSNMLKGFAPAAAAQAVQTAAQNGVRAMSSILGSSSLGGSGVGVAANLGRAA	313
Qy	241	SV 342	
Dd	314	SV 315	

RESULT 10
US-09-072-596-102
Sequence 102, Application US/09072596
Patent No. 6458366
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond

Qy	121	TAARNQLMNNVPOALKOLAAQQTQGTTPSSKLGGLWKTIVSPHRSPISNNVSMANNHSMNTN	180
Db	335	TAARNQLMNNVPOALQQLAAQTQGTTPSSKLGGLWKTIVSPHRSPISNNVSMANNHSMNTN	394
Qy	181	SGVSMNTNTLSMLKGFAPAAAQAQVOTAAQNGVFRAMSSLGSSILGSSGLGGGVAANILGRAA	240
Db	395	SGVSMNTNTLSMLKGFAPAAAQAQVOTAAQNGVFRAMSSLGSSILGSSGLGGGVAANILGRAA	454
Qy	241	SV 242	
Db	455	SV 456	

```

RESULT 15
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/8-8,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tr1-fusion
US-09-287-849-2

```

Search completed: June 22, 2004, 17:27:31
Job time : 9.38833 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 24.3589 Seconds

(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-12

Perfect score: 1306

Sequence: 1 VAWMSVTAGQBELTAQVRV.....YGHRRDGGKXYSRRNGGPA 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

-Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PTCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PTCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1304	99.8	263	12	US-09-886-349A-12
2	1304	99.8	263	14	US-10-193-002-92
3	1304	99.8	263	14	US-10-084-843-91
4	1304	99.8	263	14	US-10-098-732A-12
5	1304	99.8	358	9	US-09-287-849-8
6	1304	99.8	358	14	US-10-359-460-8
7	1187	90.9	391	12	US-09-886-349A-14
8	1187	90.9	391	14	US-10-193-002-102
9	1187	90.9	391	14	US-10-084-843-107
10	1187	90.9	391	14	US-10-098-732A-14
11	1187	90.9	596	9	US-09-287-849-26
12	1187	90.9	596	12	US-09-886-349A-20
13	1187	90.9	596	14	US-10-359-460-26
14	1187	90.9	596	14	US-10-098-732A-20
15	1187	90.9	600	9	US-09-287-849-22

16	1187	90.9	600	14	US-10-359-460-22	Sequence 22, Appl
17	1187	90.9	723	15	US-10-369-983-2	Sequence 2, Appl
18	1187	90.9	723	12	US-09-886-349A-18	Sequence 18, Appl
19	1187	90.9	723	14	US-10-098-732A-18	Sequence 18, Appl
20	1187	90.9	729	15	US-10-369-983-21	Sequence 21, Appl
21	1187	90.9	729	15	US-10-369-983-22	Sequence 22, Appl
22	1187	90.9	813	15	US-10-369-983-15	Sequence 15, Appl
23	1187	90.9	825	15	US-10-369-983-14	Sequence 14, Appl
24	1187	90.9	875	15	US-10-369-983-13	Sequence 13, Appl
25	1187	90.9	930	14	US-10-098-732A-65	Sequence 65, Appl
26	1187	90.9	930	15	US-10-369-983-12	Sequence 12, Appl
27	1187	90.9	1010	15	US-10-369-983-4	Sequence 4, Appl
28	1187	90.9	1016	15	US-10-369-983-18	Sequence 18, Appl
29	1187	90.9	1022	15	US-10-369-983-17	Sequence 17, Appl
30	1187	90.9	1154	15	US-10-369-983-16	Sequence 16, Appl
31	1182	90.5	729	9	US-09-287-849-2	Sequence 2, Appl
32	1182	90.5	729	12	US-09-886-349A-16	Sequence 16, Appl
33	1182	90.5	729	14	US-10-359-460-2	Sequence 2, Appl
34	1182	90.5	729	14	US-10-098-732A-16	Sequence 16, Appl
35	1182	90.5	729	15	US-10-359-459-2	Sequence 2, Appl
36	1179	90.3	391	12	US-09-872-186-8	Sequence 106, App
37	1001.5	76.7	396	14	US-10-193-002-106	Sequence 111, App
38	1001.5	76.7	396	14	US-10-084-843-111	Sequence 104, App
39	949.5	72.7	359	14	US-10-193-002-104	Sequence 109, App
40	949.5	72.7	359	14	US-10-084-843-109	Sequence 62455, A
41	949.5	72.7	393	12	US-10-282-122A-62455	Sequence 64892, A
42	949.5	72.7	393	12	US-10-282-122A-64892	Sequence 126, App
43	429.5	32.9	400	9	US-09-073-009-126	Sequence 126, App
44	429.5	32.9	400	9	US-09-793-306-126	Sequence 62027, A
45	429.5	32.9	405	12	US-10-282-122A-62027	

ALIGNMENTS

RESULT 1

US-09-886-349A-12
; Sequence 12, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 12
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTP39 (TbH9)
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (254)
; OTHER INFORMATION: Xaa = any amino acid
US-09-886-349A-12

Query Match 99.8%; Score 1304; DB 12; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.1e-106;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VAWMSVTAGQBELTAQVRVAAARAYATAYGLTVPPVIAENRAELMILIAIATNLGQNTPA 60

Db 1 VAWMSVTAGQBELTAQVRVAAARAYATAYGLTVPPVIAENRAELMILIAIATNLGQNTPA 60

; SEQ ID NO 14
 ; LENGTH: 391
 ; TYPE: PRT
 ; ORGANISM: Mycobacterium tuberculosis
 ; FEATURE:
 ; OTHER INFORMATION: Mtb39 full length (TbH9FL)
 US-10-098-732A-14

Query Match 90.9%; Score 1187; DB 14; Length 391;
 Best Local Similarity 99.6%; Pred. No. 3.8e-96;
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMILIAATNLLGQNTPA 60
 DB 74 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMILIAATNLLGQNTPA 133
 QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLLEQAAVEASD 120
 DB 134 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLLEQAAVEASD 193
 QY 121 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPISNNVSMANNHMSMTN 180
 DB 194 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPISNNVSMANNHMSMTN 253
 QY 181 SGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLSGSSGLGGVAAANLGRAA 240
 DB 254 SGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLSGSSGLGGVAAANLGRAA 313
 QY 241 SV 242
 DB 314 SV 315

RESULT 11
 US-09-287-849-26
 ; Sequence 26, Application US/09287849
 ; Patent No. US2002009459A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
 ; FILE REFERENCE: 014058-009020US
 ; CURRENT APPLICATION NUMBER: US/09/287,849
 ; CURRENT FILING DATE: 1999-04-07
 ; PRIOR APPLICATION NUMBER: US 08/818,112
 ; PRIOR FILING DATE: 1997-03-13
 ; PRIOR APPLICATION NUMBER: US 08/942,578
 ; PRIOR FILING DATE: 1997-10-01
 ; PRIOR APPLICATION NUMBER: US 09/025,197
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 09/056,556
 ; PRIOR FILING DATE: 1998-04-07
 ; PRIOR APPLICATION NUMBER: US 09/223,040
 ; PRIOR FILING DATE: 1998-12-30
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: Patent in Ver. 2.1
 ; SEQ ID NO 26
 ; LENGTH: 596
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: bi-fusion
 US-09-287-849-26

Query Match 90.9%; Score 1187; DB 9; Length 596;
 Best Local Similarity 99.6%; Pred. No. 6.6e-96;
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMILIAATNLLGQNTPA 60
 DB 82 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMILIAATNLLGQNTPA 141
 QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLLEQAAVEASD 120
 DB 142 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLLEQAAVEASD 201
 QY 121 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPISNNVSMANNHMSMTN 180
 DB 202 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPISNNVSMANNHMSMTN 261
 QY 181 SGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLSGSSGLGGVAAANLGRAA 240
 DB 262 SGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLSGSSGLGGVAAANLGRAA 321
 QY 241 SV 242
 DB 322 SV 323

RESULT 12
 US-09-886-349A-20
 ; Sequence 20, Application US/09886349A
 ; Publication No. US20040086523A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009070US
 ; CURRENT APPLICATION NUMBER: US/09/886,349A
 ; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: US 09/597,796
 ; PRIOR FILING DATE: 2000-06-20
 ; PRIOR APPLICATION NUMBER: US 60/265,737
 ; PRIOR FILING DATE: 2001-02-01
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: Patent in Ver. 2.1
 ; SEQ ID NO 20
 ; LENGTH: 596
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: bi-fusion
 ; OTHER INFORMATION: protein TbH9-Ra35 (designated MTB59F)
 US-09-886-349A-20

Query Match 90.9%; Score 1187; DB 12; Length 596;
 Best Local Similarity 99.6%; Pred. No. 6.6e-96;
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMILIAATNLLGQNTPA 60
 DB 82 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMILIAATNLLGQNTPA 141
 QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLLEQAAVEASD 120
 DB 142 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGLLLEQAAVEASD 201
 QY 121 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPISNNVSMANNHMSMTN 180
 DB 202 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPISNNVSMANNHMSMTN 261
 QY 181 SGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLSGSSGLGGVAAANLGRAA 240
 DB 262 SGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLSGSSGLGGVAAANLGRAA 321
 QY 241 SV 242
 DB 322 SV 323

RESULT 13
 US-10-359-460-26
 ; Sequence 26, Application US/10359460
 ; Publication No. US2003014791A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
 ; TITLE OF INVENTION: and Their Uses
 ; FILE REFERENCE: 014058-009020US
 ; CURRENT APPLICATION NUMBER: US/10/359,460
 ; CURRENT FILING DATE: 2003-02-05
 ; PRIOR APPLICATION NUMBER: US/09/287,849
 ; PRIOR FILING DATE: 1999-04-07
 ; PRIOR APPLICATION NUMBER: US 08/818,112
 ; PRIOR FILING DATE: 1997-03-13
 ; PRIOR APPLICATION NUMBER: US 08/942,578
 ; PRIOR FILING DATE: 1997-10-01
 ; PRIOR APPLICATION NUMBER: US 09/025,197
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 09/056,556
 ; PRIOR FILING DATE: 1998-04-07
 ; PRIOR APPLICATION NUMBER: US 09/223,040
 ; PRIOR FILING DATE: 1998-12-30
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 26
 ; LENGTH: 596
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
 US-10-359-460-26
 Query Match 90.9%; Score 1187; DB 14; Length 596;
 Best Local Similarity 99.6%; Pred.No. 6.6e-96;
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAMSVTAGQAEHTAAQVRVAAAAAYETAYGLTVPVPIAENRAELMILIATNLLGQNTPA 60
 Db 82 VAMSVTAGQAEHTAAQVRVAAAAAYETAYGLTVPVPIAENRAELMILIATNLLGQNTPA 141
 QY 61 IAVNEAEYGEWQAQDAAMFGYAAATATATATATLLPPEEAPMTSAGLLEQAAAVEEASD 120
 Db 142 IAVNEAEYGEWQAQDAAMFGYAAATATATATLLPPEEAPMTSAGLLEQAAAVEEASD 201
 QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMYSMANNHSMTN 180
 Db 202 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMYSMANNHSMTN 261
 QY 181 SGVSMNTNTLSSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSGLGGVAAANLGRAA 240
 Db 262 SGVSMNTNTLSSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSGLGGVAAANLGRAA 321
 QY 241 SV 242
 Db 322 SV 323
 RESULT 14
 US-10-098-732A-20
 ; Sequence 20, Application US/10098732A
 ; Publication No. US2003017529A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Brannon, Mark
 ; APPLICANT: Guderian, Jeffrey
 ; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
 ; TITLE OF INVENTION: Leishmania Antigen
 ; FILE REFERENCE: 014058-012010US
 ; CURRENT APPLICATION NUMBER: US/10/098,732A
 ; CURRENT FILING DATE: 2003-04-29
 ; PRIOR APPLICATION NUMBER: US 60/275,837
 ; PRIOR FILING DATE: 2001-03-13
 ; NUMBER OF SEQ ID NOS: 80
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 20
 ; LENGTH: 596
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
 ; OTHER INFORMATION: protein TBH9-Ra35 (designated MTB59F)
 US-10-098-732A-20
 Query Match 90.9%; Score 1187; DB 14; Length 596;
 Best Local Similarity 99.6%; Pred.No. 6.6e-96;
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAMSVTAGQAEHTAAQVRVAAAAAYETAYGLTVPVPIAENRAELMILIATNLLGQNTPA 60
 Db 82 VAMSVTAGQAEHTAAQVRVAAAAAYETAYGLTVPVPIAENRAELMILIATNLLGQNTPA 141
 QY 61 IAVNEAEYGEWQAQDAAMFGYAAATATATATLLPPEEAPMTSAGLLEQAAAVEEASD 120
 Db 142 IAVNEAEYGEWQAQDAAMFGYAAATATATATLLPPEEAPMTSAGLLEQAAAVEEASD 201
 QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMYSMANNHSMTN 180
 Db 202 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMYSMANNHSMTN 261
 QY 181 SGVSMNTNTLSSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSGLGGVAAANLGRAA 240
 Db 262 SGVSMNTNTLSSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSGLGGVAAANLGRAA 321
 QY 241 SV 242
 Db 322 SV 323
 RESULT 15
 US-09-287-849-22
 ; Sequence 22, Application US/09287849
 ; Patent No. US20020009459A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
 ; TITLE OF INVENTION: and Their Uses
 ; FILE REFERENCE: 014058-009020US
 ; CURRENT APPLICATION NUMBER: US/09/287,849
 ; CURRENT FILING DATE: 1999-04-07
 ; PRIOR APPLICATION NUMBER: US 08/818,112
 ; PRIOR FILING DATE: 1997-03-13
 ; PRIOR APPLICATION NUMBER: US 08/942,578
 ; PRIOR FILING DATE: 1997-10-01
 ; PRIOR APPLICATION NUMBER: US 09/025,197
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 09/056,556
 ; PRIOR FILING DATE: 1998-04-07
 ; PRIOR APPLICATION NUMBER: US 09/223,040
 ; PRIOR FILING DATE: 1998-12-30
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 22
 ; LENGTH: 600

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-22

Query Match 90.9%; Score 1187; DB 9; Length 600;
Best Local Similarity 99.6%; Pred. No. 6.6e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAITNLLGQNTPA 60
Db |||||
QY 82 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAITNLLGQNTPA 141
Db |||||
QY 61 IAVNEAEYGENWAGDAAMFGYAAATATATATATLLPFEEAPEMTSAGLLEQAAAVEEASD 120
Db |||||
QY 142 IAVNEAEYGENWAGDAAMFGYAAATATATATATLLPFEEAPEMTSAGLLEQAAAVEEASD 201
Db |||||
QY 121 TAAANCLMNNVPOALKOLAQPTQGTTPSSKLGGLWKTVPSPHSPISNWVSMANNHMSMTN 180
Db |||||
QY 202 TAAANCLMNNVPOALKOLAQPTQGTTPSSKLGGLWKTVPSPHSPISNWVSMANNHMSMTN 261
Db |||||
QY 181 SGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSSSLGSSGLGGVAAANLGRAA 240
Db |||||
QY 262 SGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSSSLGSSGLGGVAAANLGRAA 321
Db |||||
QY 241 SV 242
Db ||
QY 322 SV 323
Db ||

Search completed: June 22, 2004, 18:07:54
Job time : 24.3569 secs

QM protein - protein search, using sw model

Run on: June 22, 2004, 16:59:04 ; Search time 7.35842 Seconds

Run On:	Date	Time	Location	Cell Updates/sec
	June 22,	2001	3438.018 Million cell updates/sec (without alignments)	

Title: US-09-886-349A-12

Perfect score:

Sequence: 1 VAWMSVTAGQAELTAAQVRV.....YGHRDGGKYAXSGRRNGGPA 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Minimum	DB seq	length: 0
Maximum	DB seq	length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Not a copy of
PTD 79.*

Database : PLK / 8 : # 1 . n i x 1 : #

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

2: 0112: 4

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1187	90.9	391	2	H70608	probable PPE prote
2	1001.5	76.7	396	2	H70741	probable PPE prote
3	949.5	72.7	393	2	G70568	probable PPE prote
4	460	35.2	403	2	H70931	probable PPE prote
5	441	33.8	423	2	G70931	probable PPE prote
6	432.5	32.1	333	2	G70929	probable PPE prote
7	424.5	32.5	421	2	H87056	PPE-family protein
8	422	32.3	409	2	A70532	probable PPE prote
9	403	30.9	413	2	F70560	probable PPE prote
10	402	30.8	391	2	B70625	probable PPE prote
11	402	30.8	463	2	G70931	probable PPE prote
12	395	30.2	408	2	G70925	probable PPE prote
13	389	29.8	394	2	G70881	probable PPE prote
14	386.5	29.6	468	2	F70932	probable PPE prote
15	379	29.0	380	2	A70646	probable PPE prote
16	375	28.7	385	2	H70503	probable PPE prote
17	352	27.0	402	2	A70882	probable PPE prote
18	348.5	26.7	365	2	F70929	probable PPE prote
19	346.5	26.5	394	2	A70504	probable PPE prote
20	346	26.5	350	2	H70929	probable PPE prote
21	329	25.2	391	2	G70922	probable PPE prote
22	328	25.1	406	2	E70675	probable PPE prote
23	325.5	24.9	423	2	G70582	probable PPE prote
24	320.5	24.3	330	2	D70575	probable PPE prote
25	317	24.3	391	2	A70663	probable PPE prote
26	282	21.6	963	2	B70524	probable PPE prote
27	272	20.8	580	2	G70570	probable PPE prote
28	271.5	20.8	3716	2	F70969	probable PPE prote
29	266	20.4	180	2	G70932	probable PPE prote

30	257	19.7	618	2	H70552	probable PPE	prote
31	255.5	19.6	678	2	A70687	probable PPE	prote
32	250.5	19.2	1053	2	B70987	probable PPE	prote
33	245	18.9	615	2	H70663	probable PPE	prote
34	243.5	18.6	443	2	C70780	probable PPE	prote
35	243	18.6	487	2	C70830	probable PPE	prote
36	242.5	18.6	346	2	H70874	probable PPE	prote
37	242	18.5	2523	2	F70846	probable PPE	prote
38	239	18.3	655	2	A70931	probable PPE	prote
39	238	18.2	3157	2	B70969	probable PPE	prote
40	236	18.1	479	2	D70676	probable PPE	prote
41	235	18.0	204	2	T45168	probable PPE	prote
42	231.5	17.7	987	2	E70808	probable PPE	prote
43	229	17.5	1436	2	B70520	probable PPE	prote
44	227	17.4	582	2	F70675	probable PPE	prote
45	223.5	17.1	645	2	F70825	probable PPE	prote

ALIGNMENTS

RESULT 1
B70608
probable ppe protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70608
R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
C: Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID:98295987; PMID:9634230
A: Accession: B70608
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-391 <COL>
A: Cross-references: GB:Z93777; GB:AU123456; NID:g3261726; PID:e311073;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: ppe

	Query Match	90.3%;	Score 1187;	DB 2;	Length 391;
	Best Local Similarity	99.8%;	Pred. No. 8.6e-74;		
	Matches 241;	Conservative	1;	Mismatches 0;	Indels 0;
Qy	1	VAMSVTAGQAE	LTAAQVRVAAAAYETAYGLTVP	PIAENRAELMILITATNLGQNTPA	60
Db	74	VAMSVTAGQAE	LTAAQVRVAAAAYETAYGLTVP	PIAENRAELMILITATNLGQNTPA	133
Qy	61	IAYNEA	EYGEWQAQDAAMEGYAAATATATAT	LLPREEAPEMTSAGLLPQAAVEASD	120
Db	134	IAYNEA	EYGEWQAQDAAMEGYAAATATATAT	LLPREEAPEMTSAGLLPQAAVEASD	193
Qy	121	TAAANQLMN	NPQALQAOQTGTPSSKLGGLWKT	VSPhRSPISNNVSMANNHMSMTN	180
Db	194	TAAANQLMN	NPQALQAOQTGTPSSKLGGLWKT	VSPhRSPISNNVSMANNHMSMTN	253
Qy	181	SGVYSMTN	TLSSMLKGFPAAAAQAVQTAQNGV	RAMSSILGSSILGSSGGVAAANLGRAA	240
Db	254	SGVYSMTN	TLSSMLKGFPAAAAQAVQTAQNGV	RAMSSILGSSILGSSGGVAAANLGRAA	313
Qy	241	SV	242		
Db	314	SV	315		

RESULT 2
H70741
probable pPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence revision 17-Jul-1998 #text change 22-Oct-1999

C:Accession: H70741
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: H70741
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-393 <COL>
 A: Cross-references: GB:Z75555; GB:AL123456; NID:G3261608; PIDN:CAA99966.1; PID:e250360;
 A: Experimental source: strain H37RV
 C: Genetics:
 A: Gene: PPE

Query Match 76.7%; Score 1001.5; DB 2; Length 396;
 Best Local Similarity 84.1%; Pred. No. 3.9e-61;
 Matches 207; Conservative 13; Mismatches 21; Indels 5; Gaps 2;
 QY 1 VAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITATNLGQNTPA 60
 DB 74 VAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITATNLGQNTPA 133
 QY 61 IAVNEAEYGEWMAQDAAMFGVAAATATATATLTPFEAPEMTSAGGLLEQAAVEEASD 120
 DB 134 IAVNEAEYGEWMAQDAAMFGVAAATATATATLTPFEAPEMTSAGGLLEQAAVEEASD 193
 QY 121 TAAANOLMNNVPOALKQLAQPTGGTTPSSKLGGLWKTVSPHRSPISNNVSMANNHSMWN 180
 DB 194 TAAANOLMNNVPOALKQLAQPTGGTTPSSKLGGLWKTVSPHRSPISNNVSMANNHSMWN 253
 QY 181 SGVSMNTLSSMLKGFAPAAAQVOTAAQNGVRAMS---LGSSLGSSGLGGGVAANL 236
 DB 254 SGVSMNTLSSMLKGFAPAAAQVOTAAQNGVRAMS---LGSSLGSSGLGGGVAANL 312
 QY 237 GRAASV 242
 DB 313 GRAASV 318

RESULT 3
 C70568
 Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C: Species: Mycobacterium tuberculosis
 C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: C70568
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-393 <COL>
 A: Cross-references: GB:Z95390; GB:AL123456; NID:G3261766; PIDN:CAB08702.1; PID:e316074;
 A: Experimental source: strain H37RV
 C: Genetics:
 A: Gene: PPE

Query Match 72.7%; Score 949.5; DB 2; Length 393;
 Best Local Similarity 79.7%; Pred. No. 1.4e-57;
 Matches 196; Conservative 14; Mismatches 31; Indels 5; Gaps 2;
 QY 1 VAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITATNLGQNTPA 60
 DB 74 VAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITATNLGQNTPA 133
 QY 61 IAVNEAEYGEWMAQDAAMFGVAAATATATATLTPFEAPEMTSAGGLLEQAAVEEASD 120

DB 134 IEANQAYSQWQDAENMYGATATATATALLPFEDAPLTNPGLLEQAAVEEAD 193
 QY 121 TAAANOLMNNVPOALKQLAQPTGGTTPSSKLGGLWKTVSPHRSPISNNVSMANNHSMWN 180
 DB 194 TAAANOLMNNVPOALKQLAQPTGGTTPSSKLGGLWKTVSPHRSPISNNVSMANNHSMWN 253
 QY 181 SGVSMNTLSSMLKGFAPAAAQVOTAAQNGVRAMS---LGSSLGSSGLGGGVAANL 236
 DB 254 TGVSMNTLSSMLKGFAPAAAQVOTAAQNGVRAMS---LGSSLGSSGLGGGVAANL 312
 QY 237 GRAASV 242
 DB 313 GRAASV 318

RESULT 4
 H70931
 Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C: Species: Mycobacterium tuberculosis
 C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C: Accession: H70931
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: H70931
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-403 <COL>
 A: Cross-references: GB:AL022021; GB:AL123456; NID:G3250699; PIDN:CAAL7728.1; PID:e1254616
 A: Experimental source: strain H37RV
 C: Genetics:
 A: Gene: PPE

Query Match 35.2%; Score 460; DB 2; Length 403;
 Best Local Similarity 43.3%; Pred. No. 3.4e-24;
 Matches 117; Conservative 36; Mismatches 73; Indels 44; Gaps 8;
 QY 1 VAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITATNLGQNTPA 60
 DB 77 VAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITATNLGQNTPA 136
 QY 61 IAVNEAEYGEWMAQDAAMFGVAAATATATLTPFEAPEMTSAGGLLEQAAVEEASD 120
 DB 137 IAVNEAEYGEWMAQDAAMFGVAAATATATLTPFEAPEMTSAGGLLEQAAVEEASD 195
 QY 121 TAAA-----NOLMNNVPOALKQLAQPTGGTTPSSKLGGLWKTVSPHRSPISNNVSMANNHSMWN 161
 DB 196 ASAGAQQQTLLSQLIAAIPSVLQGLSSSTAATFASGPSGLLGIVGSSWLDKLLWLLDPN 255
 QY 162 RSPISNNVSMANNHSMWN---TNLTS---SMLKGFAPAAAQVOTAAQNGVRAMS 217
 DB 256 -----SNFWNTIASSGLFPISNTIAPFLGLGVAADAAGDVLSGTS----- 300
 QY 218 SLGSSL-----GSSGLGGGVAANLGRAASV 242
 DB 301 GLGGALVAPLGSAGGLGCTVAAGLGNAAV 330

RESULT 5
 B70931
 Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C: Species: Mycobacterium tuberculosis
 C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C: Accession: B70931
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; PMID:98295987; PMID:9634230
A:Accession: B70931
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-423 <COL>
A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAA17722.1; PID:e125461
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: PPE

Query Match 33.8%; Score 441; DB 2; Length 423;
Best Local Similarity 41.2%; Pred. No. 7e-23;
Matches 117; Conservative 30; Mismatches 67; Indels 70; Gaps 9;
QY 1 VAMSVTAGAQLTAAQVRVAAAYETAYGLTVPPVIAENRAELMLLIATNLLGQNTPA 60
DB 72 VAMSVATVHAELAGAARLAAYEAFAATPPVPIAANRAQLMVLIAITNIFGQNTPA 131
QY 61 IAVNEAEYGEWGAQDAAMFGYAAATATATATLPPPEAPEMTSAGGLLEQAAAVEASD 120
DB 132 IMTEAQYMEWGAQDAAMFYAGSSATA-SRMTAFTEPQTTHGQLGAQSSAVACTAA 190
QY 121 TAAAN-----QLMNVVPAKOLAQPT---QGTFP-----SSKLGGLWKTVP 160
DB 191 TAAGNLQSAFPPLLSAVPRALQGLALPTASQASATPQWVDLGNLSTFLGG-AVTGP 248
QY 161 HRSPISNVSMNNHSMNTSGVMTNTLSMLKGFAPAPAAAQAVQTAAGVYRAMSSLG 220
DB 249 YTFP-----GVLPSPGVPYLLGIQSVL-----VTQNGQGVSAALLG 283
QY 221 S-----SLGSSGLGGG-VAANLGRAASV 242
DB 284 KIGGKPTGALAPLAEFALHTPILGSEGLGGVSAGIGRAGLV 327

RESULT 6
G70929
probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70929
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; PMID:98295987; PMID:9634230
A:Accession: G70929
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-393 <COL>
A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAA17711.1; PID:e125461
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: PPE

Query Match 33.1%; Score 432.5; DB 2; Length 393;
Best Local Similarity 41.0%; Pred. No. 2.4e-22;
Matches 107; Conservative 38; Mismatches 85; Indels 31; Gaps 7;
QY 1 VAMSVTAGAQLTAAQVRVAAAYETAYGLTVPPVIAENRAELMLLIATNLLGQNTPA 60
DB 73 VAMVSAQAQAQAAQAAQAAAAFAAFAATVPPLIAANRASLMQLISTNVFGQNTSA 132
QY 61 IAVNEAEYGEWGAQDAAMFGYAAATATATLPPPEAPEMTSAGGLLEQAAAVEASD 120
DB 133 IAAAEAYGEWGAQDSAMVAYAGSSASASA-VTFSTPQIANPQAQGTQAAAVATAG 191
QY 121 TAAAN-NOLMNVVPAKOLAQPTQGTTPSSKLGGLWK-----TVSPH 161

DB 192 TAQSTILTEMITGLPNALQSLTSPLLQSS-NGPLSLMWQLFGTFNPTTSISALLTDLQPY 250
QY 162 RSPINSMVSMANNHSMNTSGVMTNTLSMLKGFAPAPAAAQAVQTAAGVYRAMSSLSGS 221
DB 251 ASFFYNTEGLPYFSIGMGNFTQSARKTL-GLIGSAAPAAVA-----AAGDAAKGLPGLGG 304
QY 222 SLGSSGLGGVAAANLGRAASV 242
DB 305 MLG-----GGPVAAGLGNAAASV 321

RESULT 7
H87056
PPE-family protein [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: H87056
R: Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holroyd, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, S.; Seeger, K.; Skelton, S.; Squares, S.; Taylor, K.; Whitehead, S.; Barrell, B.G.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squares, R.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; PMID:21128732; PMID:11234002
A:Accession: H87056
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-421 <STO>
A:Cross-references: GB:AL450380; NID:g13093150; PIDN:CAC31563.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML1182

Query Match 32.5%; Score 424.5; DB 2; Length 421;
Best Local Similarity 38.8%; Pred. No. 9.3e-22;
Matches 104; Conservative 36; Mismatches 99; Indels 29; Gaps 5;
QY 3 WMSVTAGQBELTAAQVRVAAAYETAYGLTVPPVIAENRAELMLLIATNLLGQNTPA 62
DB 75 WLTONAASAEALTATQLTVAANAYETAFTMTVPPLMVFVNRAQACLLIMSNIFGQNSTA 134
QY 63 VNEAEYGEWGAQDAAMFGYAAATATATLPPPEAPEMTSAGGL----- 108
DB 135 EKEAYETENWIDAAAMTSYQASVLEAVGATKATAPLGVNEVGLAEVVEEVEV 194
QY 109 --LEQAAAVEASD--TAAANQLMNN-----VPQALJKQ---LAQPTQGTTPSSKLG--L 154
DB 195 EVVEEVEAEQAISQAALDQAVNEGWEATVPQVDQVNVVATPQTAVPDSSSAAAPQL 254
QY 155 WKTVPSPHRSPISNVSMANNHSMNTSGVMTNTLSMLKGFAPAPAAAQAVQTAAGVYR 214
DB 255 WGGFAQHUSPINDTLTSLMNNHAGMANAGLSLVNGMSGAMKSLAP-TTTKAAESAPKAMGS 313
QY 215 AMSSLGSSILGSSGLGGVAAANLGRAASV 242
DB 314 AVCSTGRGLLGLSSSGGHVTAQLGRAASI 341

RESULT 8
A70932
probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70932
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; PMID:98295987; PMID:9634230
A:Accession: A70932
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

Query Match	30.8%;	Score 402;	DB 2;	Length 463;
Best Local Similarity	35.7%;	Pred. No. 3.6e-20;		
Matches 114;	Conservative 37;	Mismatches 106;	Indels 62;	Gaps 8;
Qy	1	VAMSVTAGQAEELTAAGVRYVAAAAYETAYGTVTPPVIAENRAELMLIIATNLLGQTPA	60	
Db	72	VAMVSAATLAAREAAQASAAAAAYEAAATATVPVPPVVAANRAELAVLAATNIGQNTGA	13.	
Qy	61	IAYNEAEYGEWMAQDAAMFCYAAATATATATLLPFEEAPEMTSAGGLLEQAAAAVEASD	120	
Db	132	IAAAEARYAEMWQDAAMCYAGSSSVAT-QVTPFPFAAPPTTAAAGLATQGVAVQAQVG	190	
Qy	121	TAAANQLMNNVPQALKOLAQETQGTTPSSKUGLWKTVS--PHRSPI-----SN	167	
Db	191	ASAGN-ARSLVSEVLEFLA--TAGTNYNKTVASLMNAVTVGPYASVYVNSMLGLGF AESK	247	
Qy	168	MVSMANNHM-----SMTNSGVSMVTNLTSSWLKGFAPAA	200	
Db	248	MVLPAQDVTIVSTIFGMVQFQKFFNPVTFNPDLIPKSGALGAGLGRSAISSGLGSTAPAI	307	
Qy	201	AAQAVQTAQAQNGVR-----AMBSLGSLSGSGGGVAAANLGRA-----ASV	242	
Db	308	SAGASAGSVGGMSPVPSWAAATPAIRTVAAVFSTGLQAVPAAI SEGSLSLQWALLASV	367	
Qy	243	RYHRDGGKYAXSGRRNGG	261	
Db	368	AGGALGGAAARATGGFTGG	386	

RESULT 12
G70925.
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70925
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:9825987; PMID:9634230
A:Accession: G70925
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-408 <COL>
A:Cross-references: GB:274024; GB:AL123456; NID:G3250700; PID:CAA98377.1; PID:el301025;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match	30.2%;	Score 395;	DB 2;	Length 408;
Best Local Similarity	38.9%;	Pred.No.9.2e-20;		
Matches 110;	Conservative 31;	Mismatches 78;	Indels 64;	Gaps 11;
Qy	1	VAMSVTAGOAELTAAQVRVAAAAYETAYGTTVPPVTAENRAELMILIAINLLGONTPA	60	
Db	73	VAMLSATAGQAEQAGMQRARAAAAYELAFANTVPPVVVAENALLVALVATINFGONTPA	132	
Qy	61	IAVNEASYGENWAOADAAAMFGYAAATATATATATLLPPEAPEMTSAGGLLEQAAA	115	
Db	133	IAATEAQYAEWMOADAAAMYAYGSAAIAT-ELTPFTAAPYTTSPAALAGQAAATVSSTV	191	
Qy	116	EEASDTAAANCLMNN-----VP--QALKQ--LAOPTGTTTSSKL-----GGL	154	
Db	192	PPLATTAAVQLLOQLSSTSLIPHYSALOQWLAENLLGLTPDNEMTVIRLLGISYFDEGL	251	
Qy	155	WKTVSPHRSPISNNMVSWANNHMTNGSVSTNTLSSMLKGFAFAAAQAQVOTAAQNGVR	214	
Db	252	-----LQFFASTLAQQAIPTCTPGAG--DSGSSVLDWSGPTIFA-----GPR	290	
Qy	215	AMSSL--GSSLGS-----SGLGGGVAANLGRAAS	241	

[illegible]

Query Match	29.8%;	Score	389;	DB 2;	Length	394;	
Best Local Similarity	37.6%;	Pred.	No.	2.3e-19;			
Matches	103;	Conservative	43;	Mismatches	106;	Gaps	7;
Qy	1	VAMSVTAQQAELTAAQVRVAAAAAYTAYGLTVPPPIVAENRAELMILIATNLLGNTTPA	60				
Db	73	LAWLTVTAESAHAHQMASAAAFAEAFAMTVPPAEVAANRALLAALVATNVLGQNTPA	132				
Qy	61	IAVNEASYGEMWADAAAMFGYAAATAATATATLLPEEEAPENTMSAGILLEQAAAAVEASD	120				
Db	133	IMATEAHYGEMWQDALAMGYGAASSA-AAGRUNLPITPSQTANMAGLQAQAAVSHHAA	191				
Qy	121	TAAANC-----LMNNVPQALKQAQPTQGTTPSSKLGGLWKTVSP--HRSPISNMYSMAN	173				
Db	192	ASTVQQVGSLGISLNPNVAMGFASPULTSAADAAGLGGIIQDEELLGITTFVQNAINNAV	251				
Qy	174	NHMSMTNSGVSMNTLTSSMLKGFPAPAAAQAVOTAAQNGV---RANSSSGSLGSSLGG	230				
Db	252	N-----TTAFVVMATINAVFLGHAFALNPATVTAAADVAPAAAAAGLAHTVTFPVGVGG	307				
Qy	231	-GVAANLGRAASVRYGHRDGGKYAXSGRRNGGPA	263				
Db	308	ASLTASILGEASSV-----GGLSVPAWGWTAPA	335				

RESULT 14

B70932

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C;Accession: B70932

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: B70932

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-468 <COL>

A;Cross-references: GB:AL020201; GB:AL123456; NID:G3250699; PTDN:CAAL17730.1; PTD:el25462

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: PPE

Job time : 8.35842 secs

```

Query Match      29.6%; Score 386.5; DB 2; Length 468;
Best Local Similarity 37.8%; Pred. No. 4.1e-19;
Matches 102; Conservative 37; Mismatches 84; Indels 47; Gaps 7;

QY      1 VAWMSVTAGQAELTAAQVRVAAAYETAYGLTVPPIAENRAELMILIATNLLQGNTPA 60
      73 VTMGTSQAQAKAADQARAADVAYETAFAAVVPPIAANRSQLISLVATNIFQNTAA 132
QY      61 IAVNPAEYGEWMAQDAAMFYAAATATATATILLPEEAPEMTSAGGLEQAAVEEAS- 119
      133 IAAEAEYGEWMAQDTWAMFYASSATA-SRUTPTAPPQTTPNPSGLAQAAATGQATA 191
QY      120 -----DTAAANQLMNNVPQALKQAQTOGQTPSSKLGSL-----W 155
      192 LASGTNAVTTTALSAAQAFPDIIPTLLQGLA--TLSTQVTLMGQLINAI FGPTGATTY 249
QY      156 KTVSPHRSFPI SNMYSMANHSMNTNSGVSMNTLSSMLKGFAPARAQAVOTAAQNGVRA 215
      250 QNVFVTAANVTAKFSIWANDAMSAENLWTEKFVF-----WQPPPAPEIPK----- 294
QY      216 MSSLGSSILGSGLGCGVAAANI GRAASVRYG 245
      295 -SSLGAGL---GLRSGLSAGLAHAASAGLG 320

```

RESULT 15

A70646
probable pPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70646
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Cole, S.R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Connor, R.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Rajandream, M.A.; Rogers, J.; Sanger, E.; Whitehead, S.; Barrell, B.G.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70646
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-380 <COL>
A:Cross-references: GB:I283867; GB:AL123456; NID:G3261695; PIDN:CAB06278.1; PID:e291015;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: pPE

```

Query Match      29.0%; Score 379; DB 2: Length 380;
Best Local Similarity 36.0%; Pred. No. le-18;
Matches 108; Conservative 40; Mismatches 96; Indels 56; Gaps 10;

QY      1 VAWMSVTAQAELETAQVRVAAAAYETAYGLTVPPVIAENRAELMILIATNLLGQNTPA 60
DB      73 IGLVYTTAEKTCQTATQARAAALAFQAYAMTLPVPVAAANRIQLLIATTNFFGONTAA 132

QY      61 IAVNEAEYGEWMAQDAAMAAPGYAAATATATATLLPPEEAPEMTSAGGLEQAAAYEEASD 120
DB      133 IAATEAQAEWMAQDAAMAAMYGYATASA-AAALLTPFPSPRTTNPAGLTAQAAAVSQATD 191

QY      121 TAAANQLMNNVPQALKQLAQP-----TQGT-----PSK 150
DB      192 PLSL--LIETVQALQALTI SPIDFTFLDAIFAGYATGVTDQVESFVAGTIGAESN 249

QY      151 LGL-----WKTYSPHRSPISNNVMANNHMSNMSG-----VSMNTLLSML----- 193
DB      250 LGLLVNGDENPAEVTGDFGELVTSATPGGVSASAGGAASGVNTVLASVGRANSIG 309

QY      194 -KGFAFAAAQAQVTAAGNGVRAMSSL-GSLSGSSLGG--GVAANLGRAASV--RYGHR 247
DB      310 QLSVSPFSMAAPSTRPVSAISPAGLTILPGTDVAEHGMPGVPGVFAAGRASGLVPRYGR 369

```

Search completed: June 22, 2004, 17:24:51

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:46:13 ; Search time 4.37699 Seconds
(without alignments)
3126.737 Million cell updates/sec

Title: US-09-886-349A-12

Perfect score: 1306

Sequence: 1 VAMSVTAQAEALTAQVRV.....YGHRDGKXVXSGRRNGGPA 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1001.5	76.7	396	1 YD61_MYCTU	Q1031 mycobacteri
2	402	30.8	463	1 Y012_MYCTU	O53951 mycobacteri
3	395	30.2	408	1 YS92_MYCTU	Q10813 mycobacteri
4	255.5	19.6	678	1 YF48_MYCTU	Q10778 mycobacteri
5	243.5	18.6	443	1 Y878_MYCTU	Q10540 mycobacteri
6	243	18.6	487	1 Y442_MYCTU	P42611 mycobacteri
7	194.5	14.9	463	1 Y096_MYCTU	Q10892 mycobacteri
8	191.5	14.7	408	1 SRA_MYCLE	O07297 mycobacteri
9	158.5	12.1	434	1 YU18_MYCTU	P31500 mycobacteri
10	158	12.1	435	1 YU22_MYCTU	O53268 mycobacteri
11	156	11.9	178	1 YF29_MYCTU	O06246 mycobacteri
12	152	11.6	176	1 YF25_MYCTU	O50703 mycobacteri
13	114.5	8.8	1211	1 BUN2_DROME	Q24523 drosophila
14	112	8.6	2090	1 N214_HUMAN	P35658 homo sapien
15	110.5	8.5	938	1 TRG1_ECOLI	P33790 escherichia
16	107.5	8.2	354	1 YAU2_SCHPO	Q10169 schizosacch
17	106	8.1	1783	1 RAA3_CHLRE	Q3f6c4 chlamydomon
18	103.5	7.9	881	1 PRY3_YEAST	P47033 saccharomyc
19	102	7.8	1120	1 STFR_ECOLI	P76072 escherichia
20	100.5	7.7	1537	1 FLO1_YEAST	P32768 saccharomyc
21	99	7.6	329	1 PRY2_YEAST	P36110 saccharomyc
22	99	7.6	659	1 SFRG_HUMAN	O8n2m8 homo sapien
23	99	7.6	1140	1 YM96_YEAST	Q04972 saccharomyc
24	98.5	7.5	475	1 SIM1_YEAST	Q61572 mus musculu
25	98.5	7.5	553	1 FXC1_MOUSE	Q10873 mycobacteri
26	98.5	7.5	558	1 YJ83_MYCTU	P33240 homo sapien
27	98.5	7.5	577	1 CST2_HUMAN	P14859 homo sapien
28	98.5	7.5	743	1 PO21_HUMAN	P38954 saccharomyc
29	98.5	7.5	1075	1 FLO5_YEAST	Q9u136 homo sapien
30	98.5	7.5	1845	1 Z236_HUMAN	P28284 herpes simp
31	98	7.5	825	1 ICP0_HSV2H	Q01981 emericella
32	97.5	7.5	416	1 CREA_EMENI	O63850 mus musculu
33	97.5	7.5	526	1 NU62_MOUSE	

34 96.5 7.4 461 1 US45_LACLC P22865 lactococcus
35 96.5 7.4 461 1 Y514_HUMAN O60269 homo sapien
36 96.5 7.4 632 1 PO21_RAT P31503 rattus norv
37 96 7.4 939 1 SLAP_CAMPE P35827 campylobact
38 96 7.4 1025 1 SLAP_CAUCR P35828 caulobacter
39 95.5 7.3 2035 1 HFC1_HUMAN P51610 homo sapien
40 95.5 7.3 2090 1 HFC1_MESAU P51611 mesocricetu
41 95 7.3 323 1 JUND_CHICK P27921 gallus gall
42 95 7.3 774 1 STF_LABMD P03764 bacterioph
43 95 7.3 1556 1 PRO3_DROVI O9u6a1 drosophila
44 94.5 7.2 601 1 PDM1_DROME P31368 drosophila
45 94.5 7.2 780 1 CTPA_MYCLE P46839 mycobacteri

ALIGNMENTS

RESULT 1

YD61_MYCTU
ID YD61_MYCTU STANDARD; PRT; 396 AA.
AC Q11031;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rvi361c/MT1406.
GN Rvi361c OR MT1406 OR MTCY02B10.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
Olliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains".
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 275555; CAA99966.1;
DR EMBL; AE007013; AAK45669.1; --
DR PIR; H70741; H70741.
DR TIGR; MT1406;
DR InterPro; IPR000030; Microbac_PPE.

```

DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 158 159 TA -> AT (IN REF. 2).
SQ SEQUENCE 396 AA; 40015 MW; 6AF6E0D7B5F668D0 CRC64;

Query Match 76.7%; Score 1001.5; DB 1; Length 396;
Best Local Similarity 84.1%; Pred. No. 2.4e-62;
Matches 207; Conservative 13; Mismatches 21; Indels 5; Gaps 2;

QY 1 VAMSVTAGQELTAQVRAAAYETAYGLTVPPVIAENRAELMLLIATNLLGQNTPA 60
DB 74 VAMSVTAGQELTAQVRAAAYETAYGLTVPPVIAENRAELMLLIATNLLGQNTPA 133
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLLEQAAAVEASD 120
DB 134 IAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLLEQAAAVEASD 193
QY 121 TAAANQLMNNVPQALKQIAQPTQCTPSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180
DB 194 TAAANQLMNNVPQALKQIAQPTQCTPSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 253
QY 181 SGVSMATLSSMLKGFAPAAAQAVTAQAQNGVPMSS-----LGSLSGSSGLGGVAAANL 236
DB 254 SGVSMATLSSMLKGFAPAAAQAVTAQAQNGVPMSS-----LGSLSGSSGLGGVAAANL 312
QY 237 GRAASV 242
DB 313 GRAASV 318

RESULT 2
YI02 MYCTU STANDARD; PRT; 463 AA.
AC O53951;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV1802/MT1851/Mb1830.
GN RV1802 OR MT1851 OR MT049.24 OR MB1830.
OS Mycobacterium tuberculosis, and
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann J., DeBoy R., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Reddy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL022021; CAAL7723.1; -
CC EMBL; AE007044; AAK46123.1; -
CC EMBL; BX248340; CAD94533.1; -
CC EMBL; C70931; C70931.
CC TIGR; MT1851; -
CC TubercuList; RV1802; -
CC InterPro; IPR000030; Microbac_PPE.
CC Pfam; PF00823; PPE; 1.
CC KW Hypothetical protein; Complete proteome.
CC FT CONFLICT 401 401 S -> L (IN REF. 2).
CC SQ SEQUENCE 463 AA; 46021 MW; E564828BF09FA551 CRC64;

Query Match 30.8%; Score 402; DB 1; Length 463;
Best Local Similarity 35.7%; Pred. No. 8.2e-21;
Matches 114; Conservative 37; Mismatches 106; Indels 62; Gaps 8;

QY 1 VAMSVTAGQELTAQVRAAAYETAYGLTVPPVIAENRAELMLLIATNLLGQNTPA 60
DB 72 VAMSVTAGQELTAQVRAAAYETAYGLTVPPVIAENRAELMLLIATNLLGQNTPA 131
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLLEQAAAVEASD 120
DB 132 IAAAEARYAEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLLEQAAAVEASD 190
QY 121 TAAANQLMNNVPQALKQIAQPTQCTPSSKLGGLWKTVPSPHRSPI-----SN 167
DB 191 ASAGN-ARSLVSEVLEFLA--TAGTNNYKTVASLNAVTVGPVYASSVNSMLGLGFAESK 247
QY 168 MYSMANNH-----SMTNSGVSMNTLSSMLKGFAPAA 200
DB 248 MYLPANDTVISTFGWVQFQFFNPVTPFNPDLIPKSAAGLGLRSGAISGLGSTAPAI 307
QY 201 AAQAVTAQAQNGVR-----AMSLGSSLSGSSGLGGVAAANLGRA-----ASV 242
DB 308 SAGASQAGSGVSGMSVPPPSMAAATPAIRTAAVFSSTGLQAVPAAAISEGLLSQMALASV 367
QY 243 RYGHDRDGGKXVXSGRRNGG 261
DB 368 AGGALGGAAARATGGFLGG 386

RESULT 3
YI02 MYCTU STANDARD; PRT; 408 AA.
AC Q10513;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV2892c/MT2959/Mb2916c.
GN RV2892c OR MT2959 OR MT2974.23C OR MB2916c.
OS Mycobacterium tuberculosis, and
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;

```


FT CONFLICT 258 258 D -> G (IN REF. 2).
 SQ SEQUENCE 678 AA; 66736 MW; 209F1593D52533A2 CRC64;

Query Match 19.6%; Score 255.5; DB 1; Length 678;
 Best Local Similarity 27.3%; Pred. No. 1.7e-10;
 Matches 72; Conservative 34; Mismatches 97; Indels 61; Gaps 6;

QY 3 WMSVTAGAELETAQVRVAAAYETAYGLTVPPIAENRAELMILLIATNLGONTAIA 62
 DB 75 WITTHAAQAEQAQAQAQAAVSAFAALAAATVHPGAVSANRGLSLVASNLGQNAFAIA 134

QY 63 VNEAYEGEMWAQDAAMFGYAAATATATATATLPEEAPEMTSAGGLLEQAAAVEASDTA 122
 DB 135 AVEAYEQMWADVAAMGLYHGEASAVALSUTPTTSP-----SAAATPGGAVII 184

QY 123 AANQLMN--NVPQALKQLAQPTGGTPSSKLGGLKWTSPHRSPISNVNMANNHMSWTN 180
 DB 195 AGFPFLDLGV-----TIGGF-----NLASNLGLGN 211

QY 181 SGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLSGLSSGLGGVAAANL----- 236
 DB 212 LG-----SFNPGSANTGSVNLGNANIGDLNLSGNIGSVNLGGGTGDLNPD 259

QY 237 GRAASVRYCHRDGGKYXSGRRNG 260
 DB 260 GNTGTLNMGSGNIGSYNLGGNGLG 283

RESULT 5
 Y478 MYCTU STANDARD; PRT; 443 AA.
 AC Q10540;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein RV0878c/MT0901.
 GN RV0878C OR MT0901 OR MTCY31.06C.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 CC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RX NCBI_TaxID=1773;
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 EX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=42206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeSoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL J. Bacteriol. 184:5479-5490 (2002).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL; Z73101; CAA97385.1; --
 EMBL; AB006977; AAK45143.1; ALT_INIT.
 PIR; C70780; C70780.
 TIGR; MT0901; --
 DR Tuberculoist; RV0878c; --
 DR InterPro; IPR000030; Microbac_PPE.
 DR InterPro; IPR002989; Mycobac_pentapep.
 DR Pfam; PF01469; Pentapeptide_2; 4.
 DR Pfam; PF00823; PPE; 1.
 KW Hypothetical protein, Transmembrane; Repeat; Complete proteome.
 FT TRANSMEM 15 35 POTENTIAL.
 FT TRANSMEM 38 58 POTENTIAL.
 FT TRANSMEM 59 79 POTENTIAL.
 FT TRANSMEM 181 201 POTENTIAL.
 FT DOMAIN 64 73 POLY-ALA.
 FT DOMAIN 81 115 ALA-RICH.
 FT DOMAIN 231 270 4 X 10 AA APPROXIMATE REPEATS.
 FT SEQUENCE 443 AA; 43592 MW; C58BEC607F0675E2 CRC64;

Query Match 18.6%; Score 243.5; DB 1; Length 443;
 Best Local Similarity 29.3%; Pred. No. 7.2e-10;
 Matches 82; Conservative 38; Mismatches 99; Indels 61; Gaps 12;

QY 1 VAMSVTGAQAELETAQVRVAAAYETAYGLTVPPIAENRAELMILLIATNLGONTAIA 60
 DB 76 LSWLNAATARAEGNAAGAKAAAYEAARAATAHPALVAARNQLSLVLNLFQNLPA 135

QY 61 IAVNEAYEGEMWAQDAAMFGYAAATATATATLPEEAPEMTSAGGLLEQAAAVEASD 120
 DB 136 IAATEASYEQLEWAQDAAMVGVHGGASVQSQTLPWQ-----LLSVLPFVWTAAP 186

QY 121 TAAANQLMNVPQALKQLAQPTGGTPSSKLGGLKWTSPHRSPISNMV---SMANNMS 177
 DB 187 AGAV-----GVPA--LAIPALGV-----ENIG-----VGNFLGIGNIGNNVG 223

QY 178 MTNSG-----VSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLSG-----SL 223
 DB 224 SNTGDTNFGNIGNVAN--LGNGNIGNANLGSNGAGPFNFGNGDNTNFGSNAGFLNI 282

QY 224 GSSGLGGVAAANLGRAASVRYGHR--DGGKYXSGRRNGG 261
 DB 283 GSGNEGSG--NLG-----FGNAGDDNTGNGSGDNTTG 313

RESULT 6
 Y442 MYCTU STANDARD; PRT; 487 AA.
 AC P42611; O53727;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein RV0442c/MT0458.
 GN RV0442C OR MT0458 OR MT037.06C.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 CC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RX NCBI_TaxID=1773;
 RP SEQUENCE FROM N.A.
 RC STRAIN=Erdmann;
 RX MEDLINE=87137260; PubMed=3029018;
 RA Shinnick T.M.;
 RL "The 65-kilodalton antigen of Mycobacterium tuberculosis."
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;


```

RESULT 7
Y096_MYCTU
ID Y096_MYCTU STANDARD; PRt; 463 AA.
AC Q10892;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV0096/MT0105.
GN RV0096 OR MT0105 OR MTCV251.15.
GE
OS Mycobacterium tuberculosis
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1773;
[1]
RN
RS SEQUENCE FROM N.A.
RC
RC STRAIN=H37Rv;
RC MEDLINE=98295937; PubMed=96314230;
RX
RX Cde S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
[2]
RN
RS SEQUENCE FROM N.A.
RC
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleisemann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Weeks T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490 (2002).
CC
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC of the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; Z74410; CAA989932.1; -.
CC DR EMBL; AE006922; AAK4327.1; -.
CC DR PIR; H70750; H70750.
CC DR TIGR; MT0105; -.
CC DR TubercuList; RV0096; -.
CC DR InterPro; IPR000030; Microbac_PPE.
CC DR Pfam; PF08243; PPE; 1.
CC DR Hypothetical protein; Transmembrane; Complete proteome.
CC KW FT TRANSMEM 3 23 POTENTIAL.
CC FT TRANSMEM 88 108 POTENTIAL.
CC FT TRANSMEM 112 132 POTENTIAL.
CC FT TRANSMEM 216 236 POTENTIAL.
CC FT TRANSMEM 245 265 POTENTIAL.
CC FT TRANSMEM 276 295 POTENTIAL.
CC FT TRANSMEM 323 343 POTENTIAL.
CC FT TRANSMEM 419 439 POTENTIAL.
CC SQ SEQUENCE 463 AA; 46894 MW; 42D9D66A033D0DD8 CRC64;
Query Match 14.9%; Score 194.5; DB 1; Length 463;
Best Local Similarity 36.3%;
Pred. No. 1.8e-06;

```


RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL J. Bacteriol. 184:5479-5490(2002).
RN [3]
RP SEQUENCE OF 160-374 FROM N.A.
RC STRAIN=Isolate 50410;
RA Patki A.H., Dale J.W.;
RL Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -!- CAUTION: In strain Oshkosh the gene for this protein is
CC interrupted in position 307 by an IS6110 element.
CC -!- CAUTION: Was originally (Ref.3) thought to be a dihydrofolate
CC reductase.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 294; 337 and 355.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL021287; CAA16103.1; -;
DR EMBL; AB007129; AAK47427.1; ALT_SEQ.
DR EMBL; X59271; CAA41961.1; ALT_SEQ.
DR PIR; E70857; E70857.
DR TIGR; MT3109; -;
DR TIGR; MT3101; -;
DR TubercuList; RV3018c; -;
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 434 AA; 43029 MW; 41D673C4BD389DD6 CRC64;

Query Match 12.1%; Score 158.5; DB 1; Length 434;
Best Local Similarity 23.4%; Pred. No. 0.00053;
Matches 67; Conservative 34; Mismatches 130; Indels 55; Gaps 7;

QY 1 VAMSVTAGQAELETAQVRVAAAYETAYGLTVPFPVIAENRAELMILITNLLGQNTPA 60
DB 76 VAMLVQASADSAAAAGEHEAAAGYVCALAEEMPTLPDLAANHLTHAVLVATNFFGINTIP 135

QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATILPPEAPEMTSAGLLEQAAAVEASD 120
DB 136 IALNEADYVRVMVQAATVMSAYEAVVGAALVATPHTGPAPVIVKPG-----ANEASN 187

QY 121 TAAAN-----QLMNVVPAQLKLAQPQGTTTPSSKGLGKMTKTSFPH---RSPIS 166
DB 188 AVAAATITPPWHEIVQFLSEETAAVDYDYLALLSELPA--VAVWVQLFVDILGFIIG 245

QY 167 NVYSMANNHSMNVSGVS-----MTNLTSSMLKGA 197
DB 246 FIITLASNAGLLTEFAINAGSYAVGLLYATAGVIDIVVEWVIGNLPGWVPLGGLL 305

QY 198 PAAQAQAVQ-TAAQNGVRMSSLSGLSGS-SGLGGGVAAALGRAAS 241
DB 306 AAADVFGVAGLAGVAGLAALPAVGAAGAPAAALVGSVAPVSGGVVS 351

RESULT 10
YU21 MYCTU
ID YU21 MYCTU STANDARD; PRT; 435 AA.
AC O53268; O53269;

DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV3021C/RV3022C/MT3106.
GN RV3021C/RV3022C OR MT3106 OR MTV012.35C/MTV012.36C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22208494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 82.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL021287; CAA16106.1; ALT_FRAME.
DR EMBL; AL021287; CAA16107.1; ALT_FRAME.
DR EMBL; AB007129; AAK47435.1; -;
DR TIGR; MT3106; -;
DR TubercuList; RV3021C; -;
DR TubercuList; RV3022C; -;
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR Hypothetical protein; Complete proteome.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 299 299 G -> A (IN REF. 2).
FT CONFLICT 317 320 LAGV -> VTGL (IN REF. 2).
FT CONFLICT 326 326 L -> V (IN REF. 2).
SQ SEQUENCE 435 AA; 42876 MW; 3B157643EAB8484A CRC64;

Query Match 12.1%; Score 158; DB 1; Length 435;
Best Local Similarity 21.9%; Pred. No. 0.00057;
Matches 67; Conservative 32; Mismatches 113; Indels 94; Gaps 7;

QY 1 VAMSVTAGQAELETAQVRVAAAYETAYGLTVPFPVIAENRAELMILITNLLGQNTPA 60
DB 76 VAMLVQASADSAAAAGEHEAAAGYVCALAEEMPTLPDLAANHLTHAVLVATNFFGINTIP 135

QY 61 IAVNEAEYGEWMAQDAAMFGYAA-----ATATATAT 92
DB 136 IALNEADYVRVMVQAATVMSAYEAVVGAALVATPHTGPAPVIVKPGANEASNAVAATIT 195

```

DR   InterPro: IPR000030; Microbac_PPE.
DR   Pfam: PF00823; PPE_1. Complete proteome.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE 178 AA; 19811 MW; 8BE1FC025ABFBEA6 CRC64;

Query Match          11.9%;   Score 156;   DB 1;   Length 178;
Best Local Similarity 37.1%;   Pred. NO. 0.0029;
Matches 39; Conservative 16; Mismatches 50; Indels 0; Gaps 0;

QY   3  WMSVTAGQBELTAQVRYVAZAAAYETAYGTVTPPPIAENRAELMLIATNLIGQNTPAIA 62
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  |||
Db    73  WUTKSRQILEYATYDFYAYVEETRHKVPFATIANRREBVRHLIASNVAGVNTPAIA 132

QY   63  VNEAEYGENWQAQAAAMFCYAAATATATATLLFPBEAPEMTSAGG 107
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  |||
Db    133  GLDAQYQOYRAQNIAMNDYQSTARFIAYLPRWOPPEOIVGCGG 177

```

```

RESULT 12
ID YY25_MYCTU STANDARD; PRT; 176 AA.
AC Q50703;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein RV3425.
DE RV3425 OR MTCY78.04C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OC NCBI_TAXID=1773;
RN [1]
SEQUENCE FROM N.A.
STRAIN=H37Rv;
MEDLINE=98295987; PubMed=96314230;
Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seesger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrall E.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z77165; CAB01031.1; -
CC PIR; F70738; F70738.
CC Tuberculist; RV3425.
CC InterPro; IPR000030; Microbac_PPE.
CC Pfam; PF00823; PPE; 1.
CC Hypothetical protein: Complete proteome.
CC SEQUENCE 176 AA; 19855 MW; B8CEP2E9463B87B0 CRC64;
CC
Query Match 11.6%; Score 152; DB 1; Length 176;
Best Local Similarity 36.5%; Pred. No. 0.00054;
Matches 38; Conservative 17; Mismatches 49; Indels 0; Gaps 0;
QY 3 WMSVTGAQELTAQVRVAAAYATGLTVPVPVIAENRAELMILTATLLGQNTPAIA 62
DDB 73 WLSKSSQLKFAAWVINGLANAYNDTRKVVPEIIPAANREERRRLTASNAGVNTPAIA 132
QV 63 VNEAEYEMWAQDAAMFGVAAATATATATATLLPPEAPEMTSAG 106

```


Db 828 TAQQ---AAAAAATSAVTPPPQNTSNAAV---TTGGQQTMLLSHMTSVEQQPN 881
 QY 71 MWAQDAAMFGYAAATATATATLIP---FEAPMTSAGLLEQAAAVEASDTAAANQL 127
 Db 882 LCAAAAAGGAAGTAATSAVAPQAFTLQLOQAP-----STIADPQOL 923
 QY 128 MNVPFQAL-----KOLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMWSNA--NNHMS 178
 Db 924 M--VPOQQQQQHQEQQQQPPQ---LPPANIASANSNINL 966
 QY 179 TN-----SGVMTNTL-----SSMLKGPAPAAAQAV-QTAA-----QNGV 213
 Db 967 TTNVATGEATNTALTITDEQATAALAAAFATGAAAAATGATSAATAATQCOIQLOQP 1026
 QY 214 RAMSSGLSSGL 228
 Db 1027 NAESETESAGTSV 1041

RESULT 14

N214 HUMAN
 ID N214 HUMAN STANDARD; PRT; 2090 AA.
 AC P35658;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Nuclear pore complex protein Nup214 (Nucleoporin Nup214) (214 kDa nucleoporin) (CAN protein).
 DB NUP214 OR CAN OR CAIN.
 GN NUP214 OR CAN OR CAIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_TaxID=9606;
 RX MEDLINE=9219315; PubMed=1549122;
 RA Von Lindern M., Fornerod M., Van Baal S., Jaegle M., De Wit T., Buijs A., Grosveld G.;
 RT "The translocation (6;9), associated with a specific subtype of acute myeloid leukemia, results in the fusion of two genes, dek and can, and the expression of a chimeric, leukemia-specific dek-can mRNA.";
 RL Mol. Cell. Biol. 12:1667-1697(1992).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=94151361; PubMed=8108440;
 RA Kraemer D., Wozniak R.W., Blobel G., Radu A.;
 RT "The human CAN protein, a putative oncogene product associated with myeloid leukemogenesis, is a nuclear pore complex protein that faces the cytoplasm.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:1519-1523(1994).
 RL
 CC -!- FUNCTION: May serve as a docking site in the receptor-mediated import of substrates across the nuclear pore complex.
 CC -!- SUBUNIT: Homodimer. Interacts with DDX19 and NUP88.
 CC -!- SUBCELLULAR LOCATION: Nuclear pore complex. Cytoplasmic filaments.
 CC -!- TISSUE SPECIFICITY: Expressed in thymus, spleen, bone marrow, kidney, brain and testis, but hardly in all other tissues or in whole embryos during development.
 CC -!- DOMAIN: Contains F-G repeats.
 CC -!- PTM: Probably glycosylated as it reacts with wheat germ agglutinin (WGA).
 CC -!- DISEASE: Implicated in a subset of acute myeloid leukemia (acute nonlymphocytic leukemia) (AML) carrying a chromosomal translocation t(6;9)(p23;q34) that results in the formation of a DEK-CAN fusion gene.
 CC -!- DISEASE: Involved in some cases of acute undifferentiated leukemia (AUL) through a chromosomal translocation t(6;9)(q21;q34.1) that involves NUP214/CAN and SET.
 CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.; WWW="http://www.infobiogen.fr/services/chronocancer/Genes/CAN.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X64228; CAA45535.1; -;
 DR PIR; S26058; S26058;
 DR Genew; HGNC:8064; NUP214.
 DR MIM; 114350; -;
 DR GO; GO:0005643; C:nuclear pore; TAS.
 DR GO; GO:0005215; F:transporter activity; TAS.
 DR InterPro; IPR004325; Nucleoporin_FG.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF03093; Nucleoporin_FG; 22.
 DR SMART; SM00320; WD40; 2.
 DR Nuclear protein; Transport; Proto-oncogene; Chromosomal translocation;
 KW Repeat; Glycoprotein.
 FT DOMAIN 481 2076 11 X 5 AA APPROXIMATE REPEATS.
 FT DOMAIN 1409 2084 18 X 4 AA APPROXIMATE REPEATS.
 FT DOMAIN 1427 2085 11 X 3 AA APPROXIMATE REPEATS.
 FT DOMAIN 1213 2090 PRO/SER/THR-RICH.
 FT DOMAIN 740 768 LEUCINE-ZIPPER 1.
 FT DOMAIN 861 882 LEUCINE-ZIPPER 2.
 FT SITE 812 813 BREAKPOINT.
 SQ SEQUENCE 2090 AA; 213766 MW; 6DBE767FDD857F8F CRC64;
 Query Match 8.6%; Score 112; DB 1; Length 2090;
 Best Local Similarity 23.1%; Pred. No. 5;
 Matches 71; Conservative 37; Mismatches 130; Indels 70; Gaps 13;
 QY 9 QCAELTAQVRVA-----AAAYET---AYGLTVPFVI 38
 Db 1592 GQTAVTAATTAAGPVAETSTPIASTTTSIVAFGPSAAAFGTVTSGSSVFPAA 1651
 QY 39 AENRAELMLLTATNLGONTAIAVNEAEYEMQAQDAAMFGYAAATATATLIP--- 95
 Db 1652 SSSS-----AFNQLTNTATAPSPVFCQVAATAPSLFGQQTGSTASTAAATQVS 1704
 QY 96 ---FEAPMTSAGLLEQA-----AAVEASDTAAANQLMNVP--QALKOLAQFTQGT 146
 Db 1705 SSGFSSPAFGTAPGVQQTTFGQASVFGQSSAASVSPQCFSSVPAGGQPA-SST 1763
 QY 147 PSSKLGGLWKTVPSPHRSPISNMWSMANNHMTNSGVSMNTLSMLKGFAP----- 198
 Db 1764 PTSTSGSVFGAASSTSS--SSSFSG--QSSPNTGGGLFGQSNAPAFGOSPGFGQGSV 1818
 QY 199 -AAAAAQVQTAQNGVRAMSSLSGLSSGLGG--GVAANLGRAASVRYGHRDGGKYXS 255
 Db 1819 FGGTSAATTTAATSGFSFCA--SGFGSNTGVSFGQAASGTGGIVFGQQSSSSSVFGS 1876
 QY 256 GR--RNGG 261
 Db 1877 GNTGRGGG 1884
 RESULT 15
 TRGL_ECOLI
 ID TRGL_ECOLI STANDARD; PRT; 938 AA.
 AC P33750;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Trag protein.
 GN Escherichia coli.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 ON NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.

```
RC STRAIN=X12;
RX MEDLINE=92204127; PubMed=1348105;
RA Firth N., Skurray R.A.;
RT "Characterization of the F plasmid bifunctional conjugation gene,
RL trag.";
RL Mol. Gen. Genet. 232:145-153(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94359430; PubMed=7915817;
RA Frost L.S., Ippen-Ihler K., Skurray R.A.;
RT "Analysis of the sequence and gene products of the transfer region of
RL the F sex factor.";
RL Microbiol. Rev. 58:162-210(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=X12 / CR63;
RA Shimizu H., Saitoh Y., Suda Y., Uehara K., Sampei G., Mizobuchi K.;
RT "Complete nucleotide sequence of the F plasmid: its implications for
RL organization and diversification of plasmid genomes.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC !- FUNCTION: PLAYS A CRUCIAL ROLE IN DONOR-RECIPIENT CELL
CC INTERACTIONS. REQUIRED FOR TWO STAGES OF THE CONJUGATION PROCESS:
CC PILUS BIOSYNTHESIS AND MATING AGGREGATE STABILIZATION. MAY
CC INTERACT WITH TRAN.
CC !- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC !- PTM: TRAG* PROBABLY ARISES FROM THE POST-TRANSLATIONAL CLEAVAGE OF
CC TRAG.
CC !- CAUTION: TRAG IS NOT RESPONSIBLE FOR THE N-TERMINAL ACETYLATION
CC OF F PILIN AS STATED BY SOME AUTHORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M59763; AAA98081.1; -.
DR EMBL; U01159; AAC44184.1; -.
DR EMBL; AP001918; BAA97969.1; -.
DR PIR; S20480; S20480.
DR EcoGene; EG40114; trag.
KW Plasmid; Conjugation; Transmembrane; Inner membrane;
Complete proteome.
FT CHAIN 1 938 TRAG PROTEIN.
FT CDS ? 938 TRAG* PROTEIN.
FT DOMAIN 1 53 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 54 73 POTENTIAL.
FT DOMAIN 74 329 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 330 350 POTENTIAL.
FT DOMAIN 351 412 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 413 433 POTENTIAL.
FT DOMAIN 434 938 PERIPLASMIC (POTENTIAL).
SQ SEQUENCE 938 AA; 102471 MW; 9E09C8402ACDFDDF CRC64;

Query Match 8.5%; Score 110.5; DB 1; Length 938;
Best Local Similarity 22.8%; Pred.No.2.6;
Matches 67; Conservative 40; Mismatches 122; Indels 65; Gaps 14;

QY 11 AELTAQVRVAAAYETAYGLTVPPV-----IAENRAELMILIAITNLLGQNTPA--- 60
DB 395 SELSOLQKYSLNASTAGYLSAMIPPLSNMVKGLGAGFSSVYSHFASSISPTASAAGS 454
QY 61 IAVNEAEYGM-----WAQDAAMFGY-----AATATATATLFFEEAPEPTSA 105
DB 455 VVDGNYSYGNQTEENVNGFSWSTNSTTSFGQMYQTGSGATATQTR-----DGNMYNDAS 509
QY 106 GGLLEQAAAYEASDFAAANCLM-----NNVPOALKQAQPTQCTTPSSKLGGLWKTVS- 159
DB 510 GMSRLPVGINATRFQIAAQOEMAREASNRASALHGF-----SSSIASAWNTLSQ 560
QY 160 --PHRSPISNMVSNVNMHNSMTNSGVSMNTLSTSLMKGFAPA---AAQAVQTAANGVR 214
```

```
Db 561 FGSNRGSSDSVTGGADSTMSAQDS--MMASRMRSAYESYAKAHNI SNEQATRELGRSTN 618
QY 215 AMSLIG-----SSLGSS--GLGGVAAANLGRAASVRYCHRDGGKYAXSGRR 258
Db 619 A--SLGLYGDAYAKHGLGISVLNGGGVGLQAGAKASIDGSDILDSHE-ASSGSR 669

Search completed: June 22, 2004, 17:16:53
Job time : 7.4679 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:55:54 ; Search time 21.5043 Seconds
(without alignments)
3858.816 Million cell updates/sec

Title: US-09-886-349a-12

Perfect score: 1306

Sequence: 1 VAMWSVTAGQAEATAQVRV.....YGHEDGGKYASGRRNGGFA 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriapi:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1187	90.9	391	16	O05298
2	1158.5	88.7	390	16	Q7U0E9
3	1122.5	85.9	396	16	Q7U071
4	1008.5	77.2	393	16	Q8VIZ3
5	949.5	72.7	393	16	O06341
6	949.5	72.7	393	16	Q7TWFS
7	463	35.5	399	16	Q7TZH8
8	460	35.2	403	16	O53956
9	441	33.8	423	16	O53950
10	441	33.8	423	16	Q7TZI4
11	432.5	33.1	393	16	Q7TZJ3
12	432.5	33.1	411	16	O53939
13	424.5	32.5	421	16	Q925K0
14	422	32.3	409	16	O53957
15	420	32.2	409	16	Q7TZH7
16	420	32.2	410	2	Q99Q11

17	403	30.9	413	16	O06386
18	402	30.8	391	16	P96362
19	402	30.8	391	16	Q7U075
20	390	29.9	394	16	Q7TXX5
21	389	29.8	462	16	O33310
22	387.5	29.7	694	16	O8VJW0
23	386.5	29.6	468	16	O53958
24	379	29.0	380	16	P95190
25	379	29.0	380	16	Q7TX66
26	375	28.7	385	16	O33204
27	375	28.7	385	16	Q8VJZ0
28	375	28.7	385	16	Q7TZR7
29	361	27.6	381	16	Q7TX67
30	355	27.2	382	16	Q7TXX3
31	355	27.2	397	2	Q9AGF0
32	352	27.0	402	16	O33312
33	348.5	26.7	365	16	O86373
34	347.5	26.6	405	16	Q8VJW5
35	347	26.6	364	16	Q7TZJ5
36	346.5	26.5	394	16	O33205
37	346.5	26.5	394	16	Q7TZB6
38	346	26.5	350	16	Q7TZJ2
39	346	26.5	363	16	O53940
40	339.5	26.0	443	16	Q8VKL9
41	333.5	25.5	443	16	Q7U242
42	332.5	25.5	187	16	Q7TZH6
43	329	25.2	391	16	O05798
44	329	25.2	391	16	Q7TX76
45	328	25.1	406	16	P71869

ALIGNMENTS

RESULT 1

O05298 ID O05298 PRELIMINARY; PRT; 391 AA.

AC O05298;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein (PPE family protein).
GN RV1196 OR MTG1364.08 OR MT1234.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Bishai W., Biglieri K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jorgels K., Krogan A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey O.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the ENBL/GenBank/DBJ databases.

O06386	mycobacteri
P96362	mycobacteri
Q7U075	mycobacteri
Q7TXX5	mycobacteri
O33310	mycobacteri
Q8VJW0	mycobacteri
O53958	mycobacteri
P95190	mycobacteri
Q7TX66	mycobacteri
O33204	mycobacteri
Q8VJZ0	mycobacteri
Q7TZR7	mycobacteri
Q7TX67	mycobacteri
Q7TXX3	mycobacteri
Q9AGF0	mycobacteri
O33312	mycobacteri
O86373	mycobacteri
Q8VJW5	mycobacteri
Q7TZJ5	mycobacteri
O33205	mycobacteri
Q7TZB6	mycobacteri
Q7TZJ2	mycobacteri
O53940	mycobacteri
Q8VKL9	mycobacteri
Q7U242	mycobacteri
Q7TZH6	mycobacteri
O05798	mycobacteri
Q7TX76	mycobacteri
P71869	mycobacteri

DR EMBL; Z93777; CAB07839.1; -.
 DR EMBL; AE007000; AAK45491.1; -.
 DR PIR; B70608; B70608.
 DR TIGR; MT1234; -.
 DR TubercuList; Rv1196; -.
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 391 AA; 39158 MW; E409396B3ABDC0F8 CRC64;

Query Match 90.9%; Score 1187; DB 16; Length 391;
 Best Local Similarity 99.6%; Pred. No. 4.5e-73;
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
 DB 74 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 133

QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGLLLEQAAAVEASD 120
 DB 134 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEEAPEMTSAGLLLEQAAAVEASD 193

QY 121 TAAANQLMNNVPQALKQIAQPTQGTTPSSKLGGLWKTVPSPHSPTSNMVMANHHMSTMN 180
 DB 194 TAAANQLMNNVPQALKQIAQPTQGTTPSSKLGGLWKTVPSPHSPTSNMVMANHHMSTMN 253

QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSGLGGVAAANLGRAA 240
 DB 254 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSGLGGVAAANLGRAA 313

QY 241 SV 242
 DB 313 SV 314

Query Match 88.7%; Score 1158.5; DB 16; Length 390;
 Best Local Similarity 97.9%; Pred. No. 3.9e-71;
 Matches 237; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
 DB 74 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 133

QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEEAPEMTSAGLLLEQAAAVEASD 120
 DB 134 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEEAPEMTSAGLLLEQAAAVEASD 193

RESULT 2

ID Q7U0E9 PRELIMINARY; PRT; 390 AA.
 AC Q7U0E9;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE PPE family protein.
 GN PPE18 OR MB1228.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL; BX248338; CAD94089.1; -.
 KW Complete proteome.
 SQ SEQUENCE 390 AA; 42786276BAB0B436 CRC64;

Query Match 88.7%; Score 1158.5; DB 16; Length 390;
 Best Local Similarity 97.9%; Pred. No. 3.9e-71;
 Matches 237; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
 DB 74 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 133

QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEEAPEMTSAGLLLEQAAAVEASD 120
 DB 134 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEEAPEMTSAGLLLEQAAAVEASD 193

RESULT 3

ID Q7U071 PRELIMINARY; PRT; 396 AA.
 AC Q7U071;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE PPE family protein.
 GN PPE19 OR MB1396C.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL; BX248338; CAD94257.1; -.
 KW Complete proteome.
 SQ SEQUENCE 396 AA; 39651 MW; 698A92A50E3CBA3F CRC64;

Query Match 85.9%; Score 1122.5; DB 16; Length 396;
 Best Local Similarity 94.3%; Pred. No. 1.1e-68;
 Matches 232; Conservative 5; Mismatches 4; Indels 5; Gaps 2;

QY 1 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
 DB 74 VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 133

QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEEAPEMTSAGLLLEQAAAVEASD 120
 DB 134 IAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEEAPEMTSAGLLLEQAAAVEASD 193

QY 121 TAAANQLMNNVPQALKQIAQPTQGTTPSSKLGGLWKTVPSPHSPTSNMVMANHHMSTMN 180
 DB 194 TAAANQLMNNVPQALKQIAQPTQGTTPSSKLGGLWKTVPSPHSPTSNMVMANHHMSTMN 253

QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSGLGGVAAANL 236
 DB 254 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSGLGGVAAANL 312

QY 237 GRAASV 242
 DB 313 GRAASV 318

RESULT 4

ID Q8VIZ3 PRELIMINARY; PRT; 393 AA.
 AC Q8VIZ3;
 DT 01-WAR-2002 (TrEMBLrel. 20, Created)
 DT 01-WAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

RA PPE family protein.
GN MT3582.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; A3007161; AAK47941.1; -
DR TIGR; MT3582; -
DR GO; GO:0016020; C-membrane; IEA.
DR GO; GO:0008233; F1-peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000030; Microbac_PPE.
DR InterPro; IPR000508; Peptidase_S26.
DR Pfam; PF00823; PPE; 1.
DR PROSITE; PS00501; SPASE_I_1; 1.
SQ SEQUENCE 393 AA; 39688 MW; 86FCB67798855511 CRC64;

Query Match 77.2%; Score 1008.5; DB 16; Length 393;
Best Local Similarity 84.6%; Pred. No. 6.e-61;
Matches 208; Conservative 14; Mismatches 19; Indels 5; Gaps 2;

QY 1 VAWMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMILITATNLGQNTPA 60
DB 74 VAWMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMILITATNLGQNTPA 133

QY 61 IAVNEAEGEMWAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLLEQAAVEEASD 120
DB 134 IAVNEAEGEMWAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLLEQAAVEEASD 193

QY 121 TAAANQLMNNVPOALQKLAQPTGTTTSSKLGGLWKTVPSPISNMVSNVNNHMSMTN 180
DB 194 TAAANQLMNNVPOALQKLAQPTGTTTSSKLGGLWKTVPSPISNMVSNVNNHMSMTN 253

QY 181 SGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSS---LGSSLSGSSGLGGVAAANL 236
DB 254 SGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSS---LGSSLSGSSGLGGVAAANL 312

QY 237 GRAASV 242
DB 313 GRAASV 318

RESULT 5
ID O06341 PRELIMINARY; PRT; 393 AA.
AC O06341;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein RV3478.
GN RV3478 OR MTCY13E12.31.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
DR EMBL; Z95390; CAB08702.1; -
DR FIR; C70568; C70568.
DR TubercuList; RV3478; -
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 72.7%; Score 949.5; DB 16; Length 393;
Best Local Similarity 79.7%; Pred. No. 7e-57;
Matches 196; Conservative 14; Mismatches 31; Indels 5; Gaps 2;

QY 1 VAWMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMILITATNLGQNTPA 60
DB 74 VAWMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMILITATNLGQNTPA 133

QY 61 IAVNEAEGEMWAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLLEQAAVEEASD 120
DB 134 IAVNEAEGEMWAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLLEQAAVEEASD 193

QY 121 TAAANQLMNNVPOALQKLAQPTGTTTSSKLGGLWKTVPSPISNMVSNVNNHMSMTN 180
DB 194 TAAANQLMNNVPOALQKLAQPTGTTTSSKLGGLWKTVPSPISNMVSNVNNHMSMTN 253

QY 181 SGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSS---LGSSLSGSSGLGGVAAANL 236
DB 254 SGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSS---LGSSLSGSSGLGGVAAANL 312

QY 237 GRAASV 242
DB 313 GRAASV 318

RESULT 6
ID Q7TWFS PRELIMINARY; PRT; 393 AA.
AC Q7TWFS;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PE family protein.
DE PPE60 OR MB3505.
GN Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/57;
RX MEDLINE=22709107; PubMed=12789972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
DR EMBL; BX248346; CAD95692.1; -
KW Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 72.7%; Score 949.5; DB 16; Length 393;
Best Local Similarity 79.7%; Pred. No. 7e-57;
Matches 196; Conservative 14; Mismatches 31; Indels 5; Gaps 2;

QY 1 VAWMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMILITATNLGQNTPA 60


```
Db 284 KIGGKPTGALAPLAEFALHTPILGSEGLGGSVSAGIGRAGLV 327

RESULT 10
Q7TZ14 PRELIMINARY; PRT; 423 AA.
AC Q7TZ14;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE PPE family protein.
GN PPE29 OR MB1829.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Carrier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthey S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248340; CAD94532.1; -.
KW Complete proteome.
SQ SEQUENCE 423 AA; 41477 MW; 26E52CC271FBBF57 CRC64;

Query Match 33.8%; Score 441; DB 16; Length 423;
Best Local Similarity 41.2%; Pred. No. 3.5e-22;
Matches 117; Conservative 30; Mismatches 67; Indels 70; Gaps 9;

QY 1 VAMSVTAGQAEHTAAQVRVAAAYETAYGLTVPPVIAENRAELMILIAINLLQNTPA 60
DB 72 VAMSVATAVHAELAGQAARLAIAAYEAFAATVPPVIAANRAQLMVLIAINIFGQNTPA 131
QY 61 IAVNEAYGEMWAQDAAMFGYAAATATATATALLPFEAEPMTSAGLLLEQAAVEEASD 120
DB 132 IMWTEAQYEMWAQDAAMFGYAGSSATA-SRMTAFTEPQTTHGQLGQSSAVAQATAA 190
QY 121 TAAAN-----QLMNVVPAQLKOLAQPT-----QGTTP-----SSKLGGLWKTVP 160
DB 191 TAAGGNLQSAFFQLLSAVPRALQGLALPTASQSASATPQWVDLGNLSTFLGG--AVTGP 248
QY 161 HRSPISNMVMNHNMTNSGVMTNTLSMLKGFAPAAAQAVQTAQAQNVRAMSSLG 220
DB 249 YTFP-----GVLPSPGVYLLGIQSVL-----VTQNGQGVSAALLG 283
QY 221 S-----SLGSSGLGGG--VAANLGRAASV 242
DB 284 KIGGKPTGALAPLAEFALHTPILGSEGLGGSVSAGIGRAGLV 327

RESULT 11
Q7TZJ3 PRELIMINARY; PRT; 393 AA.
AC Q7TZJ3;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE PPE family protein.
GN PPE26 OR MB1817.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
```

```

RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin J., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248340; CAD94520.1; -.
KW Complete proteome.
SQ SEQUENCE 393 AA; 38572 MW; 41944B6E547A8AB0 CRC64;

Query Match 33.1%; Score 432.5; DB 16; Length 393;
Best Local Similarity 41.0%; Pred. No. 1.2e-21;
Matches 107; Conservative 38; Mismatches 85; Indels 31; Gaps 7;

QY 1 VAMSVTAQQAELTAQVRVAAAAYETAYGLTVPVPIAENRAELMILIAINLLGQNTPA 60
Db VAMSSAAAQAQEAATQARAAAAAFAAFAATVPPLIAANRASLMQLISTNVFGQNTSA 132
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPSEAPEMTSAGGLLEQAAAVEASD 120
Db 133 IAAEAQYGEWMAQDAAMFGYAAATATATATLLPSEAPEMTSAGGLLEQAAAVEASD 191
QY 121 TAAA--NQLMNNVPOALKQLAQQTGTPSSKLGGLWK-----TVSPH 161
Db 192 TAQSTLTEMITGLPNALQSLTSPLLQSS--NGPLSLWQLIFGTENFTPTISALLTDLQPY 250
QY 162 RSPISNMVSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAQVTAQNGVRAMSSLS 221
Db 251 ASFFYNTGELPYFSGMGNFIQAATL-GLIGSAAPAAVA-----AAGDAKGLPLGLG 304
QY 222 SLGSSGLGGVAANLGRAASV 242
Db 305 MLG-----GGPVAAGLGNRAASV 321

RESULT 12
O53939
ID O53939 PRELIMINARY; PRT; 411 AA.
AC O53939
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE PPE-family protein.
GN RV1789 OR MT1838 OR MTW049.11.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Bruch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala E.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Davlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL; AL022021; CAAL7711.1; ALT_INIT.
DR EMBL; AE007043; AAK46108.1; -.
DR PIR; G70929; G70929.
DR TIGR; MT1838; -.
DR Tuberculist; Rv1789; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF0823; PPE; 1.
KW Complete proteome.
SQ SEQUENCE 411 AA; 40557 MW; 03C90B5E0590B7DA CRC64;

Query Match 33.1%; Score 432.5; DB 16; Length 411;
Best Local Similarity 41.0%; Pred. No. 1.3e-21;
Matches 107; Conservative 38; Mismatches 85; Indels 31; Gaps 7;

QY 1 VAMSVTAQQAELTAQVRVAAAAYETAYGLTVPVPIAENRAELMILIAINLLGQNTPA 60
Db VAMSSAAAQAQEAATQARAAAAAFAAFAATVPPLIAANRASLMQLISTNVFGQNTSA 150
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPSEAPEMTSAGGLLEQAAAVEASD 120
Db 151 IAAEAQYGEWMAQDAAMFGYAAATATATATLLPSEAPEMTSAGGLLEQAAAVEASD 209
QY 121 TAAA--NQLMNNVPOALKQLAQQTGTPSSKLGGLWK-----TVSPH 161
Db 210 TAQSTLTEMITGLPNALQSLTSPLLQSS--NGPLSLWQLIFGTENFTPTISALLTDLQPY 268
QY 162 RSPISNMVSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAQVTAQNGVRAMSSLS 221
Db 269 ASFFYNTGELPYFSGMGNFIQAATL-GLIGSAAPAAVA-----AAGDAKGLPLGLG 322
QY 222 SLGSSGLGGVAANLGRAASV 242
Db 323 MLG-----GGPVAAGLGNRAASV 339

RESULT 13
Q9Z5K0
ID Q9Z5K0 PRELIMINARY; PRT; 421 AA.
AC Q9Z5K0
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Putative PPE protein (PPE-family protein).
GN ML1182 OR MLCB1701.08C.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.A., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holtroyd S., Hornsby T., Jagels K., Macleod C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RA "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
RL EMBL; AL049191; CAB39148.1; -.
DR EMBL; AL583921; CAC31563.1; -.
DR PIR; H87056; H87056.
DR Leptrona; ML1182; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF0823; PPE; 1.
KW Complete proteome.
SQ SEQUENCE 421 AA; 43119 MW; 680F5DFDCB65B4A7 CRC64;

Query Match 32.5%; Score 424.5; DB 16; Length 421;

```

```

Query Match      32.3%; Score 122; DB 16; Length 409;
Best Local Similarity 42.2%; Pred. No. 6.6e-21;
Matches 108; Conservative 34; Mismatches 92; Indels 22; Gaps 8;

1  VAMSVTAGAQLTAAQVRVAAAAAYETAYGLTVPPPVIAENRAELMILIANLLGONTPA 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
73  VAMISVTAGAEQAQAQAKAAGYETAFATVPVPIEANRALLMSLVATNIPQONTPA 132
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
61  IAVNEAEVGEWMAQDAAMFGYAAATATATATLLPFEAPMTSAGGLLEQAAAEEASD 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
133  IAATEAHYAEWMAQDAAMFGYAGSSATA-SQLAPFSEPPQTTNPSATAQSAVVAQAAG 191
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
121  TAAA-----NQLMNVPOALKOLAQPTQTTTSSKLGGLWKVTSPHRSIPSNMVSMAN 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
192  AAASDDITAOISQLISLPTLQSLA--TTATATASAG--WDTV---LQSIITILLANLT 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
174  NHMSMTNSGVY-----MTNTLSMLKFPAPAAAQAQVQTAACNGVRAMSSLT--GSSLGS-S 226
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
245  GPYSIIIGLAIPTGWNLTIFGQILGLAQNAPGVALLGPKAAAGALSLPAPURGGVIGDIT 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

227  GLGGVAAANLGRAASV 242
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
305  PLGGGATGGIARAIYV 320
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 15
Q7TZH?
Q7TZH7 PRELIMINARY; PRT: 409 AA.
AC Q7TZH7;
AD Q7TZH7;
01-OCT-2003 (TREMBLrel. 25, Created)
01-OCT-2003 (TREMBLrel. 25, Last sequence update)
01-OCT-2003 (TREMBLrel. 25, Last annotation update)
PPE family protein.
PPE32 OR MB1837.
Mycobacterium bovis.
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI taxID=1765;
[1]
SEQUENCE FROM N.A.
STRAIN=AF2122/97; PubMed=12788972;
MEDLINE=22705107;
Garnier T., Sigmeier K., Camus J.-C., Medina N., Mansoor H.,
Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
"The complete genome sequence of Mycobacterium bovis.";
Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
EMBL; BX248340; CAD94540.1; -.
Complete proteome.
SEQUENCE 409 AA; 39931 NW; D57892628B131A9E CRC64;
Query Match 32.2%; Score 420; DB 16; Length 409;
Best Local Similarity 41.6%; Pred. No. 9,1e-21;
Matches 107; Conservative 32; Mismatches 94; Indels 24; Gaps 8;
QY 1 VAWSVTGAQBELTAAQVRVAAAAYETAYGLTPPVTAENPAEIMILTIATNLLGNTPA 60
DB 73 VAWISVTGAQBAQAGAAKIAAGVYETAATVPVPIEANKALMSLVATIFGNTTFA 132
QY 61 IAVNEAEYGENWAQAAAAAFGAAATATATATLLPFEAPEMTSAGGLLEQAAAVEASD 120
DB 133 IAAATEAHYAEKWAQDAAMYGAGSSATA-SQLAPFSEPPQTTFPSATAQAQSAVQAAG 191
QY 121 TAAA-----NQLMNVPAALKQLAQPTGTTTPSSKGLGKWTYSPHRSPISNMYSMAN 173
DB 192 AAASDITAAQ-SQLISLFPSTLQSLA--TTATATASAG--WDTV---LQSITITLANLT 244
QY 174 NHMSMTNSGV---SMTNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLSGLSIGS----- 225
DB 245 GPYSIIGALPGGHWLIFGQIL-GLAQNAPGVAAALLGPKAAGALSPLAPLRGGVIADI 303
QY 226 SGLGGGVAAIUGRAASV 242

```

Wed Jun 23 16:34:26 2004

us-09-886-349a-12.rspt

Page 8

Db : |||| : ||| |
304 TELGGGATGGIARAIYV 320

Search completed: June 22, 2004, 17:22:50
Job time : 22.5043 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:44:25 ; Search time 48.0027 Seconds
(without alignments)
2301.458 Million cell updates/sec

Title: US-09-886-349A-14

Perfect score: 1949

Sequence: 1 MVDFGALPEINSARWYAGP.....SGVLVPRPPYVWPHSPAAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1949	100.0	391	2	AAW32381 Mycobacte
2	1949	100.0	391	2	AAW32449 Mycobacte
3	1949	100.0	391	2	AAW54335 Mycobacte
4	1949	100.0	391	2	AAW81702 M. tuberc
5	1949	100.0	391	2	AAW04778 Mycobacte
6	1949	100.0	391	2	AAW38989 M. tuberc
7	1949	100.0	391	2	AAW39132 M. tuberc
8	1949	100.0	391	4	AAU01888 M. tuberc
9	1949	100.0	391	5	AAE29707 Mycobacte
10	1949	100.0	391	5	AAE17571 Mycobacte
11	1949	100.0	596	2	AAW32070 Mycobacte
12	1949	100.0	596	5	AAE29710 Mycobacte
13	1949	100.0	596	5	AAE17574 Mycobacte
14	1949	100.0	599	5	AAU74599 Antigenic
15	1949	100.0	600	2	AAW32068 Mycobacte
16	1949	100.0	600	5	AAU74597 Antigenic
17	1949	100.0	723	7	ADA26354 Mycobacte
18	1949	100.0	729	4	AAO22142 Ra12-H9-3
19	1949	100.0	729	5	AAE29709 Mycobacte
20	1949	100.0	729	5	AAE17573 Mycobacte
21	1949	100.0	729	7	ADA26374 Mycobacte
22	1949	100.0	744	4	AAU01902 M. tuberc
23	1949	100.0	813	7	ADA26367 Mycobacte
24	1949	100.0	815	4	AAU01904 M. tuberc
25	1949	100.0	825	7	ADA26366 Mycobacte

26	1949	100.0	875	7	ADA26365	Mycobacte
27	1949	100.0	930	5	AAE29731	Mycobacte
28	1949	100.0	930	7	ADA26364	Mycobacte
29	1949	100.0	1010	7	ADA26356	Mycobacte
30	1949	100.0	1016	7	ADA26370	M. bovis
31	1949	100.0	1022	7	ADA26369	Mycobacte
32	1949	100.0	1154	7	ADA26368	Mycobacte
33	1945	99.8	788	4	AAU01903	M. tuberc
34	1944	99.7	394	2	AAU04779	Mycobacte
35	1944	99.7	729	5	AAE29708	Mycobacte
36	1944	99.7	729	5	AAE17572	Mycobacte
37	1939	99.5	729	7	ADA26373	Mycobacte
38	1931	99.1	729	2	AAW32059	Mycobacte
39	1902.5	97.6	726	5	AAU74588	Antigenic
40	1652.5	84.8	396	2	AAW64337	Mycobacte
41	1652.5	84.8	396	2	AAW81704	M. tuberc
42	1652.5	84.8	396	2	AAW38991	M. tuberc
43	1652.5	84.8	396	2	AAW39134	M. tuberc
44	1583	81.2	393	6	ABU36968	Protein e
45	1583	81.2	393	6	ABU34531	Protein e

ALIGNMENTS

RESULT 1
AAW32381
ID AAW32381 standard; protein; 391 AA.
XX
AC AAW32381;
XX
DT 13-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TbH-9FL.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9709429-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US014675.
XX
PR 01-SEP-1995; 95US-00523435.
PR 22-SEP-1995; 95US-00532136.
PR 22-MAR-1996; 96US-00820280.
PR 05-JUN-1996; 96US-00658800.
PR 12-JUL-1996; 96US-00680573.
XX
(CORI-) CORIXA CORP.
PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TH, Twardzik DR;
XX
DR WPI; 1997-192904/17.
XX
N-PSDB; AAT91455.
XX
New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
useful for diagnosis of M. tuberculosis infection.

Example 3; Page 150-152; 190pp; English.

A new immunogenic polypeptide has been developed comprising an immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its variant differing only in conservative substitutions and/or modifications). The present sequence represents a M. tuberculosis antigen, TbH-9FL. The immunogenic polypeptide can be used to diagnose M. tuberculosis infection by forming complexes with specific antibodies in the sample. Fragments of DNA encoding the immunogenic polypeptide can be used as diagnostic primers or probes and agents that bind to the antigen,

CC especially monoclonal antibodies or equivalent polyclonal antibodies, are
 CC also used for diagnosis

XX Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 2; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINARMYAGPGASLVAAAQWDSVADLPSAASAFQSVVWGLTVGSWIG 60
 DB 1 MVDGALPPEINARMYAGPGASLVAAAQWDSVADLPSAASAFQSVVWGLTVGSWIG 60

QY 61 SSAGLMVAASPYVAVMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
 DB 61 SSAGLMVAASPYVAVMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMI 120

QY 121 LIATNLLGQNTPTAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 180
 DB 121 LIATNLLGQNTPTAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 180

QY 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPI 240
 DB 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPI 240

QY 241 MVMANNHMTNSGVSMTNTLSSMLKGFAPAAAAQVQTAQAQNGVRAMSSLSGSSG 300
 DB 241 MVMANNHMTNSGVSMTNTLSSMLKGFAPAAAAQVQTAQAQNGVRAMSSLSGSSG 300

QY 301 LGGVVAANLGRAASVGSLSVQAWAANAQVTPAARALPLTSLTSAERPGQMLGGLPV 360
 DB 301 LGGVVAANLGRAASVGSLSVQAWAANAQVTPAARALPLTSLTSAERPGQMLGGLPV 360

QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
 DB 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391

RESULT 2

AAW32449
 ID AAW32449 standard; protein; 391 AA.

XX AAW32449;

XX 09-JAN-1998 (first entry)

XX Mycobacterium tuberculosis antigen TbH-9FL.

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 XX skin testing; M.tuberculosis.

XX Mycobacterium tuberculosis.

XX WO9709428-A2.

XX 13-MAR-1997.

XX 30-AUG-1996; 96WO-US014674.

XX 01-SEP-1995; 95US-00523436.

XX 22-SEP-1995; 95US-00533634.

XX 22-MAR-1996; 96US-00620874.

XX 05-JUN-1996; 96US-00659683.

XX 12-JUL-1996; 96US-00680574.

XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;
 XX Vedwick TH, Twardzik DR;

XX WPI; 1997-192903/17.

XX N-PSDB; AAT91521.

PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also for
 PT diagnosis.

XX Example 3; Page 138-139; 168pp; English.

XX A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
 CC variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis antigen,
 CC TbH-9FL. The immunogenic protein, and fusion proteins containing one or
 CC more of the proteins or one of the proteins plus ESAT-6, are useful in
 CC vaccines, preferably when formulated with a non-specific adjuvant, to
 CC induce an immune response against M.tuberculosis (for treatment or
 CC prevention)

XX Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 2; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINARMYAGPGASLVAAAQWDSVADLPSAASAFQSVVWGLTVGSWIG 60
 DB 1 MVDGALPPEINARMYAGPGASLVAAAQWDSVADLPSAASAFQSVVWGLTVGSWIG 60

QY 61 SSAGLMVAASPYVAVMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
 DB 61 SSAGLMVAASPYVAVMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMI 120

QY 121 LIATNLLGQNTPTAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 180
 DB 121 LIATNLLGQNTPTAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 180

QY 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPI 240
 DB 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPI 240

QY 241 MVMANNHMTNSGVSMTNTLSSMLKGFAPAAAAQVQTAQAQNGVRAMSSLSGSSG 300
 DB 241 MVMANNHMTNSGVSMTNTLSSMLKGFAPAAAAQVQTAQAQNGVRAMSSLSGSSG 300

QY 301 LGGVVAANLGRAASVGSLSVQAWAANAQVTPAARALPLTSLTSAERPGQMLGGLPV 360
 DB 301 LGGVVAANLGRAASVGSLSVQAWAANAQVTPAARALPLTSLTSAERPGQMLGGLPV 360

QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
 DB 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391

RESULT 3

AAW64335
 ID AAW64335 standard; protein; 391 AA.

XX AAW64335;

XX 17-OCT-2003 (revised)

XX 09-NOV-1998 (first entry)

XX Mycobacterium tuberculosis antigen TbH-9FL.

XX Tuberculosis; infection; diagnosis; antigen; TbH-9FL.

XX Mycobacterium tuberculosis; strain H37Rv.

XX WO9816645-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US018214.

XX 11-OCT-1996; 96US-00729622.

PR 13-MAR-1997; 97US-00818111.
XX (CORI-) CORIXA CORP.
XX
PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ;
XX
XX WPI; 1998-251292/22.
XX N-PSDB; AAV44395.
XX
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
XX develop products for the detection of M. tuberculosis infection and
XX diagnosis of tuberculosis.
XX
XX Example 3; Page 133-135; 250pp; English.
XX
XX This polypeptide comprises Mycobacterium tuberculosis antigen TbH-9FL. It
XX is encoded by genomic DNA (see AAV44395) isolated from a M. tuberculosis
XX strain H37Rv genomic library using a probe from clone TbH-9 (see
XX AAV44371). The invention relates to compositions and methods for
XX diagnosing tuberculosis. It provides polypeptides (see AAV64291-W64379)
XX comprising an antigenic portion of a soluble M. tuberculosis antigen, or
XX an immunogenic portion of an M. tuberculosis antigen, as well as DNA
XX sequences encoding such polypeptides, recombinant expression vectors and
XX transformed or transfected host cells. Also claimed are methods and
XX diagnostic kits for detecting M. tuberculosis infection in a patient
XX using these polypeptides, antibodies or oligonucleotide probes and
XX primers for the diagnosis of tuberculosis. (Updated on 17-OCT-2003 to
XX standardise OS field)
XX
XX SQ Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.2e-142;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINARMYAGPGSASLVAAQMWDSVADLFSASAFQSVVWGLTVGSGWIG 60
DB 1 MVDGALPPEINARMYAGPGSASLVAAQMWDSVADLFSASAFQSVVWGLTVGSGWIG 60
QY 61 SSAGLMVAAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTPPPVIAENRAELMI 120
DB 61 SSAGLMVAAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTPPPVIAENRAELMI 120
QY 121 LIATNLLGQNTPAIYVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLLGQNTPAIYVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MYSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAVMSISLSSLSGSG 300
DB 241 MYSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAVMSISLSSLSGSG 300
QY 301 LGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
DB 361 GOMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391

RESULT 4
AAW81702
ID AAW81702 standard; protein; 391 AA.
XX
XX AC AAW81702;
XX
XX DT 27-JAN-1999 (first entry)
XX

M. tuberculosis immunogenic polypeptide TbH-9FL.
XX
XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
XX vaccine; pharmaceutical; infection; diagnosis.
XX
XX Mycobacterium tuberculosis.
XX
XX WO9816646-A2.
XX
XX 23-APR-1998.
XX
XX 07-OCT-1997; 97WO-US018293.
XX
XX 11-OCT-1996; 96US-00730510.
XX
XX 13-MAR-1997; 97US-00818112.
XX
XX (CORI-) CORIXA CORP.
XX
XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX Vedvick TS, Twardzik DR, Lodes MJ;
XX
XX WPI; 1998-261042/23.
XX N-PSDB; AAV64503.
XX
XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
XX develop products for the detection of M. tuberculosis infection and for
XX diagnosis, treatment and prevention of tuberculosis.
XX
XX Example 3B; Page 128-129; 230pp; English.
XX
XX This sequence represents an immunogenic portion of a soluble
XX Mycobacterium tuberculosis (MT) antigen which can be used in a method for
XX inducing protective immunity against tuberculosis (TB). This sequence can
XX be formulated into vaccines and/or pharmaceutical compositions for
XX immunising against M. tuberculosis infection or may be used for the
XX diagnosis of tuberculosis
XX
XX SQ Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.2e-142;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINARMYAGPGSASLVAAQMWDSVADLFSASAFQSVVWGLTVGSGWIG 60
DB 1 MVDGALPPEINARMYAGPGSASLVAAQMWDSVADLFSASAFQSVVWGLTVGSGWIG 60
QY 61 SSAGLMVAAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTPPPVIAENRAELMI 120
DB 61 SSAGLMVAAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTPPPVIAENRAELMI 120
QY 121 LIATNLLGQNTPAIYVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLLGQNTPAIYVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MYSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAVMSISLSSLSGSG 300
DB 241 MYSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAVMSISLSSLSGSG 300
QY 301 LGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
DB 301 LGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
QY 361 GOMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
DB 361 GOMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391

RESULT 5

RAY04778
 ID AAY04778 standard; protein; 391 AA.
 AC AAY04778;
 XX
 XX
 DT 06-JUL-1999 (first entry)
 XX
 XX Mycobacterium species protein sequence 5R.
 DE
 XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.
 XX
 XX Mycobacterium sp.
 OS
 XX WO9909186-A2.
 PN
 XX 25-FEB-1999.
 PD
 XX 14-AUG-1998; 98WO-FR001813.
 PF
 XX 14-AUG-1997; 97FR-00010404.
 PR
 XX 11-SEP-1997; 97FR-00011325.
 XX
 XX (INSP) INST PASTEUR.
 FA
 XX
 XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
 PI Goguet De La Salmoniere Y;
 PI
 XX
 XX WPI; 1999-181045/15.
 DR
 XX N-PSDB; AAX34030.
 XX
 XX Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in infection-associated
 PT protein expression.
 XX
 XX Claim 32; Fig 5R; 309pp; French.
 PS
 XX
 XX Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
 CC from various Mycobacterium species microorganisms. The encoding
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection
 XX
 SQ Sequence 391 AA;
 Query Match 100.0%; Score 1949; DB 2; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142; Mismatches 0; Indels 0; Gaps 0;
 Matches 391; Conservative 0;
 QY 1 MVDFGALPPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWIG 60
 DB 1 MVDFGALPPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWIG 60
 QY 61 SSAGLMVAASPYVAMSVTAGQELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 DB 61 SSAGLMVAASPYVAMSVTAGQELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 QY 121 LIATNLGQNTPAIVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAG 180
 DB 121 LIATNLGQNTPAIVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAG 180
 QY 181 LLEQAAVEASDTAAANQLMNNVPOALQOQPTGGTTPSSKLGGLWKTVPSPHSPI 240
 DB 181 LLEQAAVEASDTAAANQLMNNVPOALQOQPTGGTTPSSKLGGLWKTVPSPHSPI 240
 QY 241 MVSVMANNHSMNTSGVSWTNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSG 300
 DB 241 MVSVMANNHSMNTSGVSWTNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSG 300
 QY 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMTLGL 360
 DB 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMTLGL 360

QY 361 GQMGARAGGGLSGVLRVPRPYPVMPHSPAAG 391
 DB 361 GQMGARAGGGLSGVLRVPRPYPVMPHSPAAG 391
 RESULT 6
 AAY38989
 ID AAY38989 standard; protein; 391 AA.
 XX
 AC AAY38989;
 XX
 XX 05-NOV-1999 (first entry)
 DT
 XX M. tuberculosis recombinant antigen protein TBH-9FL.
 DE
 XX Antigen; diagnosis; detection; infection; antibody; immunisation;
 KW vaccine; immunity.
 XX
 XX Mycobacterium tuberculosis.
 OS
 XX WO9942118-A2.
 PN
 XX 26-AUG-1999.
 PD
 XX 17-FEB-1999; 99WO-US003265.
 PF
 XX 18-FEB-1998; 98US-00024753.
 PR
 XX 05-MAY-1998; 98US-00072596.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedwick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
 PI
 XX WPI; 1999-527416/44.
 DR
 XX N-PSDB; AA219093.
 XX
 XX New polypeptide comprising antigenic portions of M. tuberculosis.
 PT
 XX Example 3; Page 168-169; 323pp; English.
 PS
 XX This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against M.
 CC tuberculosis infection. The new detection methods are needed as current
 CC vaccination strategies do not provide 100% immunity
 XX

SQ Sequence 391 AA;
 Query Match 100.0%; Score 1949; DB 2; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142; Mismatches 0; Indels 0; Gaps 0;
 Matches 391; Conservative 0;
 QY 1 MVDFGALPPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWIG 60
 DB 1 MVDFGALPPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWIG 60
 QY 61 SSAGLMVAASPYVAMSVTAGQELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 DB 61 SSAGLMVAASPYVAMSVTAGQELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 QY 121 LIATNLGQNTPAIVNEAEYGEWMAQDAAMFGYAAATATATLPPFEAPMTSAG 180
 DB 121 LIATNLGQNTPAIVNEAEYGEWMAQDAAMFGYAAATATATLPPFEAPMTSAG 180
 QY 181 LLEQAAVEASDTAAANQLMNNVPOALQOQPTGGTTPSSKLGGLWKTVPSPHSPI 240
 DB 181 LLEQAAVEASDTAAANQLMNNVPOALQOQPTGGTTPSSKLGGLWKTVPSPHSPI 240
 QY 241 MVSVMANNHSMNTSGVSWTNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSG 300

Db 241 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSSLGSSG 300
QY 301 LGGVVAANLGRAASVGSLSVPQAWAAANQAVTAAQNGVRAMSSLGSSG 360
Db 301 LGGVVAANLGRAASVGSLSVPQAWAAANQAVTAAQNGVRAMSSLGSSG 360
QY 361 GQMGARAGGSLGSLRVPRPYVMPHSPAAG 391
Db 361 GQMGARAGGSLGSLRVPRPYVMPHSPAAG 391

RESULT 7
AAV39132
ID AAV39132 standard; protein; 391 AA.
AC AAV39132;
XX
XX
DT 05-NOV-1999 (first entry)
DE M. tuberculosis antigen Tbh-9FL amino acid sequence.
DE
KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test.
XX
XX Mycobacterium tuberculosis.
XX
XX WO9942076-A2.
XX
XX 26-AUG-1999.
XX
XX 17-FEB-1999; 99WO-US003268.
XX
XX 18-FEB-1998; 98US-00025197.
PR 05-MAY-1998; 98US-00072967.
XX
XX (CORI-) CORIXA CORP.
XX
XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX
XX WPI; 1999-527405/44.
DR N-PSDB; AAZ19305.
XX
XX New antigens from Mycobacterium tuberculosis useful in diagnostic skin
PT tests and protective or therapeutic vaccines or compositions.
XX
XX Example 3; Page 123-124; 299pp; English.
XX
XX The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described are
CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
CC polypeptides fragments, can be used in pharmaceutical compositions or
CC vaccines to generate a protective or therapeutic immune response to M.
CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
CC killer cells and/or macrophages in tuberculosis-immune subjects. AAZ19249
CC to AAZ19460 and AAZ19083 to AAZ39225 are used in the exemplification of
CC the present invention
XX
XX Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.2e-142;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWIG 60
Db 1 MVDGALPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWIG 60
QY 61 SSAGLMVAASPYVAMWSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120

Db 61 SSAGLMVAASPYVAMWSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
QY 121 LIATNLGQNTPALAVNAEYGENWAOAAAFGAAAAATATATATLTPFEAPEMTSAGG 180
Db 121 LIATNLGQNTPALAVNAEYGENWAOAAAFGAAAAATATATATLTPFEAPEMTSAGG 180
QY 181 LLEQAAAVEEASDTAAANQNLMMNVPOALQLOAQTOQTTPSSKGLGWLKTVSPHRSPLN 240
Db 181 LLEQAAAVEEASDTAAANQNLMMNVPOALQLOAQTOQTTPSSKGLGWLKTVSPHRSPLN 240
QY 241 MYSVANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSSLGSSG 300
Db 241 MYSVANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSSLGSSG 300
QY 301 LGGVVAANLGRAASVGSLSVPQAWAAANQAVTAAQNGVRAMSSLGSSG 360
Db 301 LGGVVAANLGRAASVGSLSVPQAWAAANQAVTAAQNGVRAMSSLGSSG 360
QY 361 GQMGARAGGSLGSLRVPRPYVMPHSPAAG 391
Db 361 GQMGARAGGSLGSLRVPRPYVMPHSPAAG 391

RESULT 8
AAU01888
ID AAU01888 standard; protein; 391 AA.
AC AAU01888;
XX
XX 29-AUG-2001 (first entry)
DE M. tuberculosis antigen Tbh9 (Mtb39A).
XX
XX Tbh9; Mtb39A; antigen; vaccine; tuberculosis; AIDS;
KW acquired immunodeficiency disease.
XX
XX Mycobacterium tuberculosis.
XX
XX WO200124820-A1.
XX
XX 12-APR-2001.
XX
XX 10-OCT-2000; 2000WO-US028095.
XX
XX 07-OCT-1999; 99US-0158338P.
PR 07-OCT-1999; 99US-0158425P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Houghton RL, Mcneill PD, Dillon DC, Lodes ML;
PI WPI; 2001-290576/30.
XX N-PSDB; AA030779.
XX
XX Vaccinating against Mycobacteria infections in mammals using fusion
PT proteins comprising combinations of heterologous antigens.
XX
XX Example 2; Page 151-152; 168pp; English.
XX
XX The sequence represents Mycobacterium tuberculosis Tbh9 (also known as
CC Mtb39A), an M. tuberculosis antigen. Compositions comprising at least 2
CC heterologous antigens, as a fusion protein, and vectors expressing the
CC fusion proteins are used as vaccines to prophylactically immunise mammals
CC (especially humans) against infection by Mycobacteria. The compositions
CC contain at least 2 heterologous antigens that increase the serological
CC sensitivity of individuals infected with tuberculosis, a disease
CC frequently affecting patients with acquired immunodeficiency disease,
CC AIDS
XX
XX Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 4; Length 391;

Best Local Similarity 100.0%; Pred. No. 1.2e-142; Mismatches 0; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
DB 1 MYDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 60

QY 61 SSAGLWVAASPYVWMSVTAGQAELETAQVRAAAAYETAYGLTVPPPIAENRAELMI 120
DB 61 SSAGLWVAASPYVWMSVTAGQAELETAQVRAAAAYETAYGLTVPPPIAENRAELMI 120

QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLLPFEAPEMTSAGG 180

QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240

QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSG 300
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSG 300

QY 301 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAARERFGQMLGGLPV 360
DB 301 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAARERFGQMLGGLPV 360

QY 361 GQMGARAGGSLGVLRVPPRPYVMPHSPAAG 391
DB 361 GQMGARAGGSLGVLRVPPRPYVMPHSPAAG 391

RESULT 9
AAE17571
ID AAE17571 standard; protein; 391 AA.
XX
AC AAE17571;
XX
DT 27-JAN-2003 (first entry)
XX
DE Mycobacterium sp. TbH9FL antigenic protein.
XX
KW vaccine; immunity; diagnostic agent; gene therapy; TbH9FL antigen.
XX
OS Mycobacterium sp.
XX
FN WO200272792-A2.
XX
PD 19-SEP-2002.
XX
FF 13-MAR-2002; 2002WO-US008223.
XX
PR 13-MAR-2001; 2001US-0275837P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Brannon M, Guderian J;
XX
DR WPI; 2002-759844/82.
XX
DR N-PSDB; AAD47082.
XX
PT New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
XX M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
XX against pathogenic microorganisms e.g. Leishmania and Mycobacterium
XX tuberculosis.
XX
PS Disclosure; Page 85-87; 155pp; English.
XX
CC The invention relates to a recombinant nucleic acid molecule encoding a
XX fusion polypeptide. The recombinant nucleic acid comprises a heterologous
XX polynucleotide sequence encoding an antigen or an antigenic fragment from
XX Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
XX polypeptide or its fragment. The Leishmania polynucleotide is selected

CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
CC are used in methods for eliciting immune response in mammals. They are
CC useful as vaccines to elicit protective immunity against pathogenic
CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
CC polypeptides are used for enhancing the expression of polynucleotides, as
CC in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is
CC Mycobacterium sp. TbH9FL antigenic protein
XX
SQ Sequence 391 AA;
Query Match 100.0%; Score 1949; DB 5; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.2e-142; Mismatches 0; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
DB 1 MYDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 60

QY 61 SSAGLWVAASPYVWMSVTAGQAELETAQVRAAAAYETAYGLTVPPPIAENRAELMI 120
DB 61 SSAGLWVAASPYVWMSVTAGQAELETAQVRAAAAYETAYGLTVPPPIAENRAELMI 120

QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLLPFEAPEMTSAGG 180

QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240

QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSG 300
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSG 300

QY 301 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAARERFGQMLGGLPV 360
DB 301 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAARERFGQMLGGLPV 360

QY 361 GQMGARAGGSLGVLRVPPRPYVMPHSPAAG 391
DB 361 GQMGARAGGSLGVLRVPPRPYVMPHSPAAG 391

RESULT 10
AAE17571
ID AAE17571 standard; protein; 391 AA.
XX
AC AAE17571;
XX
DT 22-APR-2002 (first entry)
XX
DE Mycobacterium species MTB39 (TbH9) protein #2.
XX
KW Fusion protein; antigen; serological sensitivity; immune response;
XX tuberculosis; infection; vaccine; MTB39; TBH9 protein.
XX
OS Mycobacterium sp.
XX
FN WO200198460-A2.
XX
PD 27-DEC-2001.
XX
PF 20-JUN-2001; 2001WO-US019959.
XX
PR 20-JUN-2000; 2000US-00597796.
XX
PR 01-FEB-2001; 2001US-0265737P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Reed S, Alderson M;
XX
DR WPI; 2002-147798/19.

DR N-PSDB; AAD28341.
 XX Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 XX Claim 83; Page 102-103; 136pp; English.
 XX
 CC The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB39 (TbH9) protein
 XX
 XX Sequence 391 AA;
 SQ
 Query Match 100.0%; Score 1949; DB 5; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
 Db 1 MVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
 QY 61 SSAGLMVAASPYVAMWSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 Db 61 SSAGLMVAASPYVAMWSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 QY 121 LIATNLLGQNTPAIYVNEAEYEGEMWQDAAMFGYAAATATATATLLPFEAPENTSAGG 180
 Db 121 LIATNLLGQNTPAIYVNEAEYEGEMWQDAAMFGYAAATATATATLLPFEAPENTSAGG 180
 QY 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
 Db 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
 QY 241 MYSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQNGVRAMSSILGSSLGSSG 300
 Db 241 MYSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQNGVRAMSSILGSSLGSSG 300
 QY 301 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAABERGPGOMLGGLPV 360
 Db 301 LGGVVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAABERGPGOMLGGLPV 360
 QY 361 GQMGARAGGSLGVLVRPPRPYVMPHSPAA 391
 Db 361 GQMGARAGGSLGVLVRPPRPYVMPHSPAA 391
 RESULT 11
 AAY32070
 ID AAY32070 standard; protein; 596 AA.
 XX
 AC AAY32070;
 XX
 XX 17-JAN-2000 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen fusion protein Mtb59f.

XX Tuberculosis; antigen: fusion protein; Mtb59f; TbH9; Ra35; diagnosis;
 KW therapy; vaccine; immunogen.
 XX
 OS Mycobacterium tuberculosis.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..8
 FT Protein /note= "Met/His tag"
 FT Protein 9..140
 FT Protein /note= "Ra12"
 FT Protein 143..596
 FT Protein /note= "TbH9"
 PN WO9951748-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 07-APR-1999; 99WO-US007717.
 XX
 PR 07-APR-1998; 98US-00056556.
 PR 30-DEC-1998; 98US-00223040.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky YAW, Alderson M, Campos-Neto A;
 XX
 DR WPI; 1999-601610/51.
 DR N-PSDB; AAZ20205.
 XX
 XX New fusion proteins useful for diagnosis, prevention and treatment of
 PT tuberculosis.
 PS Claim 1; Fig 12A-B; 83pp; English.
 XX
 CC This sequence represents a recombinant Mycobacterium tuberculosis bi-
 CC antigen fusion protein, termed Mtb59f, composed of the antigens TbH9 and
 CC Ra35. The fusion protein is expressed in host cells using a vector
 CC carrying a polynucleotide (see AAZ20205) comprising the coding sequences
 CC for the 2 antigens. The invention provides fusion proteins (see AAY32059-
 CC 71) containing at least 2 M. tuberculosis antigens. The new fusion
 CC proteins and polynucleotides encoding them are useful as vaccines or
 CC preventing tuberculosis (claimed), for diagnosis (via in vitro assays or
 CC intradermal skin tests for detection of anti-M. tuberculosis antibodies),
 CC monitoring of disease progression, and treatment of tuberculosis. They
 CC are more effective immunogens than mixtures of the individual protein
 CC components
 XX
 SQ Sequence 596 AA;
 Query Match 100.0%; Score 1949; DB 2; Length 596;
 Best Local Similarity 100.0%; Pred. No. 2.1e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
 Db 9 MVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
 QY 61 SSAGLMVAASPYVAMWSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 Db 69 SSAGLMVAASPYVAMWSVTAGQAEELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
 QY 121 LIATNLLGQNTPAIYVNEAEYEGEMWQDAAMFGYAAATATATATLLPFEAPENTSAGG 180
 Db 129 LIATNLLGQNTPAIYVNEAEYEGEMWQDAAMFGYAAATATATATLLPFEAPENTSAGG 188
 QY 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
 Db 189 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 248
 QY 241 MYSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQNGVRAMSSILGSSLGSSG 300
 Db 249 MYSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQNGVRAMSSILGSSLGSSG 308

QY 301 LGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
 Db 309 LGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 368
 QY 361 GQMGARAGGGLSGVLRVPPRPVPMHSPAA 391
 Db 369 GQMGARAGGGLSGVLRVPPRPVPMHSPAA 399

RESULT 12
 AAE29710
 ID AAE29710 standard; protein; 596 AA.
 XX
 AC AAE29710;
 XX
 DT 29-AUG-2003 (revised)
 DT 27-JAN-2003 (first entry)
 XX
 DE Mycobacterium sp. MTB59F fusion protein.
 XX
 KW Vaccine; immunity; diagnostic agent; gene therapy; TbH9, antigen; Ra35;
 KW MTB59F; fusion protein.
 XX
 OS Mycobacterium sp.
 OS Mycobacterium tuberculosis.
 OS Chimeric.
 XX
 PN WO200272792-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 13-MAR-2002; 2002WO-US008223.
 XX
 PR 13-MAR-2001; 2001US-0275837P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Brannon M, Guderian J;
 XX
 DR WPI; 2002-759844/82.
 DR N-PSDB; AAD47086.
 XX
 PT New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 PT tuberculosis.
 XX
 PS Disclosure; Page 98-99; 155pp; English.
 XX
 CC The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC MTB59F fusion protein. This fusion protein comprises Ra35 protein from
 CC Mycobacterium tuberculosis and TbH9 protein from Mycobacterium sp.
 CC (Updated on 29-AUG-2003 to standardise OS field)

QY 1 MVDGALPPEINSAARMYAGFGSASLVAAAQWDSVASDLFSAASFQSVVWGLTVGSWIG 60
 Db 61 SSAGLMWAAAAPFYVAMSVTAGQAELETAQVRAAAAYETAYGLTVPPPVTAENRAELMI 120
 Db 69 SSAGLMWAAAAPFYVAMSVTAGQAELETAQVRAAAAYETAYGLTVPPPVTAENRAELMI 128
 QY 121 LIATNLGONTPTAIVNEAEYGEWMAODAAAFYAAATATATATATATLPPFEAPMTSAGG 180
 Db 129 LIATNLGONTPTAIVNEAEYGEWMAQDAANFYAAATATATATLPPFEAPMTSAGG 188
 QY 181 LLEQAAAVEEASDTAAANQLMNVFPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSISN 240
 Db 189 LLEQAAAVEEASDTAAANQLMNVFPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSISN 248
 QY 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSLSGLSSG 300
 Db 249 MVSMAHHMNTSGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSLSGLSSG 308
 QY 301 LGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
 Db 309 LGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 368
 QY 361 GQMGARAGGGLSGVLRVPPRPVPMHSPAA 391
 Db 369 GQMGARAGGGLSGVLRVPPRPVPMHSPAA 399

RESULT 13
 AAE17574
 ID AAE17574 standard; protein; 596 AA.
 XX
 AC AAE17574;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Mycobacterium species MTB59F fusion protein.
 XX
 KW Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB59F; TbH9-Ra35 protein.
 XX
 OS Mycobacterium sp.
 XX
 PN WO200198460-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 20-JUN-2001; 2001WO-US019959.
 XX
 PR 20-JUN-2000; 2000US-00597796.
 PR 01-FEB-2001; 2001US-0285737P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Reed S, Alderson M;
 DR WPI; 2002-147798/19.
 DR N-PSDB; AAD28344.
 XX
 CC Composition comprising MTB39 antigen and MTB32A antigen from
 CC Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 PS Claim 5; Page 114-115; 135pp; English.
 XX
 CC The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,

QY 301 LGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
 Db 309 LGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 368
 QY 361 GQMGARAGGGLSGVLRVPPRPVPMHSPAA 391
 Db 369 GQMGARAGGGLSGVLRVPPRPVPMHSPAA 399

Query Match 100.0%; Score 1949; DB 5; Length 596;
 Best Local Similarity 100.0%; Pred. No. 2,1e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB59F (TbH9-Ra35) fusion protein
 XX
 SQ Sequence 596 AA;

Query Match 100.0%; Score 1949; DB 5; Length 596;
 Best Local Similarity 100.0%; Pred. No. 2.1e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGIG 60
 DB 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGIG 68
 QY 61 SSAGLMVAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 DB 69 SSAGLMVAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
 DB 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 188
 QY 181 LLEQAAVEEASDTAAANQLMNVPQALQQAQPTGCTPSSKLGGLWKTVPSPHSPISN 240
 DB 189 LLEQAAVEEASDTAAANQLMNVPQALQQAQPTGCTPSSKLGGLWKTVPSPHSPISN 248
 QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSLGSSG 300
 DB 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSLGSSG 308
 QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
 DB 369 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 399

RESULT 14
 AAU74599
 AC AAU74599 standard; protein; 599 AA.
 AC AAU74599;

29-AUG-2003 (revised)
 DT 08-MAY-2002 (first entry)
 DE Antigenic fusion protein Tb59-Ra35 (Mcb59f).
 XX Fusion protein; tuberculosis; Mycobacterium tuberculosis;
 KW tuberculostatic; immunogen; vaccine; 1559-Ra35; Mcb59f.
 XX
 OS Mycobacterium tuberculosis.
 OS Chimeric.
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 597
 FT /label= OTHER
 FT /note= "OTHER= Xaa. Xaa= In frame stop codon"
 XX
 XX US2002009459-A1.
 XX 24-JAN-2002.

XX 07-APR-1999; 99US-00287849.
 XX 13-MAR-1997; 97US-00818112.
 PR 01-OCT-1997; 97US-00942578.
 PR 18-FEB-1998; 98US-00025197.
 PR 07-APR-1998; 98US-00056556.
 PR 30-DEC-1998; 98US-00223040.
 XX (REED/) REED S G.
 PA (SKEI/) SKEIKY Y A.
 PA (DILL/) DILLON D C.
 PA (ALDE/) ALDERSON M.
 PA (CAMP/) CAMPOS-NETO A.
 XX
 FI Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
 XX WPI; 2002-171134/22.
 DR N-PSDB; ABK14139.
 XX
 XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for
 PT diagnosing, treating or preventing M. tuberculosis infection,
 PT particularly as vaccine for treating or preventing tuberculosis.
 XX
 PS Claim 1; Fig 12; 62pp; English.
 XX
 CC The invention relates to a purified polypeptide which induces an immune
 CC response of Mycobacterium tuberculosis. Polypeptides of the invention are
 CC useful for diagnosing, treating or preventing M. tuberculosis infection,
 CC particularly tuberculosis infection. In particular, the polypeptides are
 CC useful as a vaccine formulation with an adjuvant to afford long-term
 CC protection in animals against the development of tuberculosis. The
 CC protein coding sequence may be used to encode a protein product for use
 CC as an immunogen to induce and/or enhance an immune response to M.
 CC tuberculosis. This sequence represents an M. tuberculosis fusion protein
 CC of the invention. (Updated on 29-AUG-2003 to standardise OS field)
 XX
 SQ Sequence 599 AA;

Query Match 100.0%; Score 1949; DB 5; Length 599;
 Best Local Similarity 100.0%; Pred. No. 2.1e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGIG 60
 DB 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGIG 68
 QY 61 SSAGLMVAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 DB 69 SSAGLMVAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
 DB 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
 QY 181 LLEQAAVEEASDTAAANQLMNVPQALQQAQPTGCTPSSKLGGLWKTVPSPHSPISN 240
 DB 189 LLEQAAVEEASDTAAANQLMNVPQALQQAQPTGCTPSSKLGGLWKTVPSPHSPISN 248
 QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSLGSSG 300
 DB 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSLGSSG 308
 QY 301 LGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABERGPGQMLGGPLV 360
 DB 309 LGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABERGPGQMLGGPLV 368
 QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
 DB 369 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 399

RESULT 15

```

AAY32068
ID AAY32068 standard; protein; 600 AA.
XX AC AAY32068;
XX DT 17-JAN-2000 (first entry)
XX DE Mycobacterium tuberculosis antigen fusion protein Mtb61f.
XX KW Tuberculosis; antigen; fusion protein; Mtb61f; TBH9; DPV; MTI; diagnosis;
XX KW therapy; vaccine; immunogen.
XX OS Mycobacterium tuberculosis.
XX PN W09951748-A2.
XX PD 14-OCT-1999.
XX PF 07-APR-1999; 99WO-US007717.
XX PR 07-APR-1998; 98US-00056556.
XX PR 30-DEC-1998; 98US-00223040.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky YAW, Alderson M, Campos-Neto A;
XX DR WPI; 1999-601610/51.
XX DR N-PSDB; AAZ20203.
XX PT New fusion proteins useful for diagnosis, prevention and treatment of
XX PT tuberculosis.
XX PS Claim 1; Fig 10A-B; 83pp; English.
XX CC This sequence represents a recombinant Mycobacterium tuberculosis tri-
CC antigen fusion protein, termed Mtb61f, composed of the antigens TbH9, Dpv
CC and MTI. The fusion protein is expressed in host cells using a vector
CC carrying a polynucleotide (see AAZ20203) comprising the coding sequences
CC for the 3 antigens. The invention provides fusion proteins (see AAY32059-
CC 71) containing at least 2 M. tuberculosis antigens. The new fusion
CC proteins and polynucleotides encoding them are useful as vaccines for
CC preventing tuberculosis (claimed), for diagnosis (via in vitro assays or
CC intradermal skin tests for detection of anti-M. tuberculosis antibodies),
CC monitoring of disease progression, and treatment of tuberculosis. They
CC are more effective immunogens than mixtures of the individual protein
CC components
XX SQ Sequence 600 AA;

Query Match 100.0%; Score 1949; DB 2; Length 600;
Best Local Similarity 100.0%; Pred. No. 2.1e-142;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPPEINGARMYAGSGASLVAAQMDSDVADLFSASAPQSVVWGLTVGSMIG 60
DB 9 MVDGALPPPEINGARMYAGSGASLVAAQMDSDVADLFSASAPQSVVWGLTVGSMIG 68
QY 61 SSAGLVAAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
DB 69 SSAGLVAAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEAPEMTSAGG 180
DB 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEAPEMTSAGG 188
QY 181 LLEQAAVEEASDPAANQLMNNVPCALQOQTGTPSSKLGGLWKTVPSPHSPISN 240
DB 189 LLEQAAVEEASDPAANQLMNNVPCALQOQTGTPSSKLGGLWKTVPSPHSPISN 248
QY 241 MYSMANNHSMNTNSGVSMNTNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSILGSSG 300
DB 249 MYSMANNHSMNTNSGVSMNTNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSILGSSG 308

```

```

QY 301 LGGGVAANLGRAASVGSLSVFPQAWAAANQAVTPAARALPLTSLTSAARFGQMLGGLPV 360
DB 309 LGGGVAANLGRAASVGSLSVFPQAWAAANQAVTPAARALPLTSLTSAARFGQMLGGLPV 368
QY 361 GOMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
DB 369 GOMGARAGGGLSGVLRVPPRPYVMPHSPAAG 399

```

Search completed: June 22, 2004, 17:15:29
Job time : 49.0027 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:07:54 ; Search time 13.9576 Seconds
(without alignments)
1446.225 Million cell updates/sec

Title: US-09-886-349A-14

Perfect score: 1949

Sequence: 1 MVDGALPPEINSARMYAGP.....SCGLRVPRPYWPHSPAAAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits: satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents RA:*

- 1: /cgn2.6/prodata2/iaa/5A_COMB.pep.*
- 2: /cgn2.6/prodata2/iaa/5B_COMB.pep.*
- 3: /cgn2.6/prodata2/iaa/6A_COMB.pep.*
- 4: /cgn2.6/prodata2/iaa/6B_COMB.pep.*
- 5: /cgn2.6/prodata2/iaa/PCFUS_COMB.pep.*
- 6: /cgn2.6/prodata2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1949	100.0	391	3	US-08-818-112-107
2	1949	100.0	391	4	US-08-818-111-102
3	1949	100.0	391	4	US-09-056-556-107
4	1949	100.0	391	4	US-09-072-967-107
5	1949	100.0	391	4	US-09-072-967-107
6	1949	100.0	596	4	US-09-287-849-26
7	1949	100.0	600	4	US-09-287-849-22
8	1944	99.7	729	4	US-09-223-040-2
9	1944	99.7	729	4	US-09-287-849-2
10	1552.5	84.8	396	3	US-08-818-112-111
11	1552.5	84.8	396	4	US-08-818-111-106
12	1552.5	84.8	396	4	US-09-056-556-111
13	1552.5	84.8	396	4	US-09-072-967-106
14	1552.5	84.8	396	4	US-09-072-967-111
15	1486.5	76.3	359	3	US-08-818-112-109
16	1486.5	76.3	359	4	US-08-818-111-104
17	1486.5	76.3	359	4	US-09-056-556-109
18	1486.5	76.3	359	4	US-09-072-967-104
19	1486.5	76.3	359	4	US-09-072-967-109
20	1187.5	60.9	358	4	US-09-287-849-8
21	1187	60.9	263	3	US-08-818-113-91
22	1187	60.9	263	4	US-08-818-111-92
23	1187	60.9	263	4	US-09-056-556-91
24	1187	60.9	263	4	US-09-072-967-92
25	1187	60.9	263	4	US-09-072-967-91
26	766.5	39.3	400	4	US-09-073-009-126
27	766.5	39.3	400	4	US-09-073-010-126

Sequence 16, Appl
Sequence 12, Appl
Sequence 142, App
Sequence 142, App
Sequence 131, App
Sequence 15, Appl
Sequence 15, Appl
Sequence 57, Appl
Sequence 208, App
Sequence 92, Appl
Sequence 114, App
Sequence 109, App
Sequence 114, App
Sequence 109, App
Sequence 114, App
Sequence 114, App
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-818-112-107
; Sequence 107, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 153
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/POCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-107

Query Match 100.0%; Score 1949; DB 3; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
Db 1 MVDGALPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
QY 61 SSAGLWVAASPPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLWVAASPPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGQNTPAIVANEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGG 180
Db 121 LIATNLLGQNTPAIVANEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGG 180
QY 181 LLEQAAVEASDTAAANQLMNNVPQALQQAQVQTAQAQVQTAQAQVQTAQAQVQTAQAQV 240
Db 181 LLEQAAVEASDTAAANQLMNNVPQALQQAQVQTAQAQVQTAQAQVQTAQAQVQTAQAQV 240
QY 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAPAAAQVQTAQAQVQTAQAQVQTAQAQVQTAQAQV 300
Db 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAPAAAQVQTAQAQVQTAQAQVQTAQAQVQTAQAQV 300
QY 301 LGGGVAANLGRAASVGSLSVQAAWAAANQVTPAARALPLTSLTSAARERGPGQMLGGLPV 360
Db 301 LGGGVAANLGRAASVGSLSVQAAWAAANQVTPAARALPLTSLTSAARERGPGQMLGGLPV 360
QY 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391
Db 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391

RESULT 2
US-08-818-111-102
; Sequence 102, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-102

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
Db 1 MVDGALPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
QY 61 SSAGLWVAASPPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLWVAASPPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGQNTPAIVANEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGG 180
Db 121 LIATNLLGQNTPAIVANEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGG 180
QY 181 LLEQAAVEASDTAAANQLMNNVPQALQQAQVQTAQAQVQTAQAQVQTAQAQVQTAQAQV 240
Db 181 LLEQAAVEASDTAAANQLMNNVPQALQQAQVQTAQAQVQTAQAQVQTAQAQVQTAQAQV 240
QY 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAPAAAQVQTAQAQVQTAQAQVQTAQAQVQTAQAQV 300
Db 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAPAAAQVQTAQAQVQTAQAQVQTAQAQVQTAQAQV 300
QY 301 LGGGVAANLGRAASVGSLSVQAAWAAANQVTPAARALPLTSLTSAARERGPGQMLGGLPV 360
Db 301 LGGGVAANLGRAASVGSLSVQAAWAAANQVTPAARALPLTSLTSAARERGPGQMLGGLPV 360
QY 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391
Db 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391

RESULT 3
US-09-056-556-107
; Sequence 107, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
TREATM

US-09-056-556-107

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 MVDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60

QY 61 SSAGLMVAASPYVAVMSVTAGQAEILTAQVRVAAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAASPYVAVMSVTAGQAEILTAQVRVAAAAAYETAYGLTVPPVIAENRAELMI 120

QY 121 LIATNLLGQNTPAIAVNEAEYGEKWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLLGQNTPAIAVNEAEYGEKWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180

QY 181 LLEQAAVAEASDTAAANLNNVPOALQOLAQPTGTTTSSKLGGLWKTVPSPHSPISN 240
DB 181 LLEQAAVAEASDTAAANLNNVPOALQOLAQPTGTTTSSKLGGLWKTVPSPHSPISN 240

QY 241 MVSMAHHMTSGVSMNTLSSMLKGFAPAAAAQAVQTAARALPLTSLTSAABRGPGQMLGGLPV 360
DB 241 MVSMAHHMTSGVSMNTLSSMLKGFAPAAAAQAVQTAARALPLTSLTSAABRGPGQMLGGLPV 360

QY 301 LGGVVAANLGRAASVGSLSVFOQAAANQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
DB 301 LGGVVAANLGRAASVGSLSVFOQAAANQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 360

QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
DB 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391

RESULT 4

US-09-072-596-102
; Sequence 102, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-102

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 MVDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60

QY 61 SSAGLMVAASPYVAVMSVTAGQAEILTAQVRVAAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAASPYVAVMSVTAGQAEILTAQVRVAAAAAYETAYGLTVPPVIAENRAELMI 120

QY 121 LIATNLLGQNTPAIAVNEAEYGEKWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLLGQNTPAIAVNEAEYGEKWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180

QY 181 LLEQAAVAEASDTAAANLNNVPOALQOLAQPTGTTTSSKLGGLWKTVPSPHSPISN 240
DB 181 LLEQAAVAEASDTAAANLNNVPOALQOLAQPTGTTTSSKLGGLWKTVPSPHSPISN 240

QY 241 MVSMAHHMTSGVSMNTLSSMLKGFAPAAAAQAVQTAARALPLTSLTSAABRGPGQMLGGLPV 360
DB 241 MVSMAHHMTSGVSMNTLSSMLKGFAPAAAAQAVQTAARALPLTSLTSAABRGPGQMLGGLPV 360

QY 301 LGGVVAANLGRAASVGSLSVFOQAAANQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
DB 301 LGGVVAANLGRAASVGSLSVFOQAAANQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 360

QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
DB 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391

RESULT 5

US-09-072-967-107
; Sequence 107, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967

```
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-967-107

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
Db 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY 61 SSAGLMVAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
Db 61 SSAGLMVAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
QY 121 LIATNLGONTPATVANEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
Db 121 LIATNLGONTPATVANEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQVQTAQNGVRAMSSLGSSLGSSG 300
Db 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQVQTAQNGVRAMSSLGSSLGSSG 300
QY 241 MVSNNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
Db 241 MVSNNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
QY 301 LGGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERFGQMLGGLPV 360
Db 301 LGGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERFGQMLGGLPV 360
QY 361 QMGARAGGGLSGVLVRPPRPYPVPHSPAAG 391
Db 361 QMGARAGGGLSGVLVRPPRPYPVPHSPAAG 391

RESULT 6
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46

; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-967-107

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
Db 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY 61 SSAGLMVAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
Db 61 SSAGLMVAASPYVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
QY 121 LIATNLGONTPATVANEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
Db 121 LIATNLGONTPATVANEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQVQTAQNGVRAMSSLGSSLGSSG 300
Db 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQVQTAQNGVRAMSSLGSSLGSSG 300
QY 241 MVSNNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
Db 241 MVSNNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
QY 301 LGGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERFGQMLGGLPV 360
Db 301 LGGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERFGQMLGGLPV 360
QY 361 QMGARAGGGLSGVLVRPPRPYPVPHSPAAG 391
Db 361 QMGARAGGGLSGVLVRPPRPYPVPHSPAAG 391

RESULT 7
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
```

Qy	61	SSAGLMVAAASP	YVAMNSVTAGQAE	LTAAQVRVAAAAAYETAYGLTVPPPIAENRAELMI	120
Db	202	SSAGLMVAAASP	YVAMNSVTAGQAE	LTAAQVRVAAAAAYETAYGLTVPPPIAENRAELMI	261
Qy	121	LIATNLLGQNT	PAIAVNEAEYEGEMW	QDAAMFGVAAATATATATALLPPEEAPEMTSAGG	180
Db	262	LIATNLLGQNT	PAIAVNEAEYEGEMW	QDAAMFGVAAATATATATALLPPEEAPEMTSAGG	321
Qy	181	LLEQAAAVEEAS	DTAAANQLMNNVPOAL	QOOLAQTGTTTSSKLGKGLWKTVSPHRSPI	240
Db	322	LLEQAAAVEEAS	DTAAANQLMNNVPOAL	QOOLAQTGTTTSSKLGKGLWKTVSPHRSPI	381
Qy	241	MVSMANNHSM	TNSGVSMNTLSSMLKGF	APAAAAQAVQTAQAQNGVRAMSSLGSSLGSSG	300
Db	382	MVSMANNHSM	TNSGVSMNTLSSMLKGF	APAAAAQAVQTAQAQNGVRAMSSLGSSLGSSG	441
Qy	301	LGGGVAANLGR	AAASVGSLSVPQAWAA	ANCAVTPAARALPLTSLTSAABERGPGQMLGGLPV	360
Db	442	LGGGVAANLGR	AAASVGSLSVPQAWAA	ANCAVTPAARALPLTSLTSAABERGPGQMLGGLPV	501
Qy	361	GQMGARAGG	LSGLVLRVPPRYM	PHSPAAG 391	
Db	502	GQMGARAGG	LSGLVLRVPPRYM	PHSPAAG 532	
<p>RESULT 9</p> <p>US-09-287-849-2</p> <p>; Sequence 2, Application US/09287849</p> <p>; Patent No. 6627198</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: Reed, Steven G.</p> <p>; APPLICANT: Skeiky, Yasir A.W.</p> <p>; APPLICANT: Dillon, Davin C.</p> <p>; APPLICANT: Alderson, Mark</p> <p>; APPLICANT: Campos-Neto, Antonio</p> <p>; APPLICANT: Corixa Corporation</p> <p>; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens</p> <p>; TITLE OF INVENTION: and Their Uses</p> <p>; FILE REFERENCE: 014058-009020US</p> <p>; CURRENT APPLICATION NUMBER: US/09/287,849</p> <p>; CURRENT FILING DATE: 1999-04-07</p> <p>; PRIOR APPLICATION NUMBER: US 08/918,112</p> <p>; PRIOR FILING DATE: 1997-03-13</p> <p>; PRIOR APPLICATION NUMBER: US 08/942,578</p> <p>; PRIOR FILING DATE: 1997-10-01</p> <p>; PRIOR APPLICATION NUMBER: US 09/025,197</p> <p>; PRIOR FILING DATE: 1998-02-18</p> <p>; PRIOR APPLICATION NUMBER: US 09/056,556</p> <p>; PRIOR FILING DATE: 1998-04-07</p> <p>; PRIOR APPLICATION NUMBER: US 09/223,040</p> <p>; PRIOR FILING DATE: 1998-12-30</p> <p>; NUMBER OF SEQ ID NOS: 46</p> <p>; SOFTWARE: PatentIn Ver. 2.1</p> <p>; SEQ ID NO 2</p> <p>; LENGTH: 729</p> <p>; TYPE: PRT</p> <p>; ORGANISM: Artificial Sequence</p> <p>; FEATURE:</p> <p>; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion</p> <p>US-09-287-849-2</p> <p>Query Match 99.7%; Score 1944; DB 4; Length 729;</p> <p>Best Local Similarity 99.7%; Pred. No. 1.8e-153;</p> <p>Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;</p>					
Qy	1	MVDFGALPPE	INSARMYAGPGSASL	VAAQMWDSVASDLFSAASAFQSVVWGLTVGSMIG	60
Db	142	MVDFGALPPE	INSARMYAGPGSASL	VAAQMWDSVASDLFSAASAFQSVVWGLTVGSMIG	201
Qy	61	SSAGLMVAAASP	YVAMNSVTAGQAE	LTAAQVRVAAAAAYETAYGLTVPPPIAENRAELMI	120
Db	202	SSAGLMVAAASP	YVAMNSVTAGQAE	LTAAQVRVAAAAAYETAYGLTVPPPIAENRAELMI	261

QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLFFBEAPMTSAGG 180
DB 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLFFBEAPMTSAGG 321
QY 181 LLEQAAVEEASDTAAANOLMNNVPOALQOIAQPTQGTTPSSKLGGLWKTVPVPHRSPISN 240
DB 322 LLEQAAVEEASDTAAANOLMNNVPOALQOIAQPTQGTTPSSKLGGLWKTVPVPHRSPISN 381
QY 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSG 300
DB 382 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSG 441
QY 301 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERPGQMLGLPV 360
DB 442 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERPGQMLGLPV 501
QY 361 QMGARAGGGLSGVLRVPPRPVYVMPHSPAAG 391
DB 502 QMGARAGGGLSGVLRVPPRPVYVMPHSPAAG 532

RESULT 10

US-08-818-112-111
; Sequence 111, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

US-08-818-112-111

Query Match 84.8%; Score 1652.5; DB 3; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.4e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

DB 1 VVDFGALPPEINARMYAGPGSASLVAAAKMWDVSASDLFSAASAFQSVVWGLTGSWIG 60
QY 61 SSAGLWVAASPYVAWMSVTAGQAEELTAAQVRAAAAYETAYGLTVPVPPVIAENRAELMI 120
DB 61 SSAGLWVAASPYVAWMSVTAGQAEELTAAQVRAAAAYETAYGLTVPVPPVIAENRAELMI 120
QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLFFBEAPMTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLFFBEAPMTSAGG 180
QY 181 LLEQAAVEEASDTAAANOLMNNVPOALQOIAQPTQGTTPSSKLGGLWKTVPVPHRSPISN 240
DB 181 LLEQAAVEEASDTAAANOLMNNVPOALQOIAQPTQGTTPSSKLGGLWKTVPVPHRSPISN 240
QY 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSG 296
DB 241 IVSMLNHNHVMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSG 299
QY 297 GSSGLGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERPGQMLG 356
DB 300 GSSGLGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERPGQMLG 359
QY 357 GLPVGMQMGARAG--GGLSGVLRVPPRPVYVMPHSPAAG 391
DB 360 GLPLQLTNSGGGFGVSNALRMPPRAYVMPRVPAAG 396

RESULT 11

US-08-818-111-106
; Sequence 106, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

US-08-818-111-106

Query Match	84.8%;	Score 1652.5;	DB 4;	Length 396;
Best Local Similarity	84.9%;	Pred. No. 1.4e-129;		
Matches 337;	Conservative 19;	Mismatches 34;	Indels 7;	Gaps 3;
QY	1	MYDFGALPPEINSRMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTGVSGWIG	60	
DB	1	VVDFGALPPEINSRMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTGVSGWIG	60	
QY	61	SSAGLWMAAASPYYVAMWSVTAGQBELTAAQVRVAAAAAYETAYGLTVPVPVIAENRAELMI	120	
DB	61	SSAGLWMAAASPYYVAMWSVTAGQBELTAAQVRVAAAAAYETAYGLTVPVPVIAENRAELMI	120	
QY	121	LIATNLLQGNTPATAVNEAEYGEWQAQDAAMFGYAAATATATATLLPPEBAPEMTSAGG	180	
DB	121	LIATNLLQGNTPATAVNEAEYGEWQAQDAAMFGYAAATATATTEALLPEDAPLTLNPGG	180	
QY	181	LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQCTTPESSKGLGKWTXPHRSPISN	240	
DB	181	LLEQAAVEEAIIDTAAANQLMNNVPQALQQLAQTQKSIWPFQDLSLWKAISPHLSPLSN	240	
QY	241	MYSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAAQAQVOTAAQNGVYRANSS---	296	
DB	241	IYSMLNNHVSMTNSGVSMASSTLHSMKLGFAF-AAAQAVETAAQNGVQAQWSSLGSGLSSL	299	
QY	297	GSSGLGGVGAANILGRAASVGSLSVFPQAWAANQAVTPAARALPLTSLTSAARRGQCMIG	356	
DB	300	GSSGLGAGVAAVLGRAASVGSLSVFPQAWAANQAVTPAARALPLTSLTSAAQTPAHMIG	359	
QY	357	GLPVGQMGARAG--CGLSGVTVRVPRPYPVPHSPAG	391	
DB	360	GLPLGQLTNSGGGFGGVSNALRMPMPRAYVMPVPPVPAAG	396	

TREATM

RESULT 12

US-09-056-556-111

; Sequence 111, Application US/09056556

; Patent No. 6350456

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND

; NUMBER OF SEQUENCES: 241

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED AND BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/056,556

; FILING DATE: 07-APR-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.457

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 111:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 396 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

US-09-056-556-111

Query Match	84.8%;	Score	1552.5;	DB	4;	Length	396;
Best Local Similarity	84.9%;	Pred. No.	1.4e-129;				
Matches	337;	Conservative	19;	Mismatches	34;	Indels	7;
Gaps	3;						
Qy	1	MYDFGALPEINSARMYAGPGSASLVAAAOQWDSVASDLFSAASAFOSVVMGLTVGSMIG	60				
Db	1	VVDFGALPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFOSVVMGLTVGSMIG	60				
Qy	61	SSAGLWVAAASPYVAVMSVTAGQABELTAAQVRYVAAAAYETAYGLTVPPPVIAENRAELMI	120				
Db	61	SSAGLWVAAASPYVAVMSVTAGQABELTAAQVRYVAAAAYETAYGLTVPPPVIAENRAELMI	120				
Qy	121	LIATNLGQNTPAIAVNEAEYGEVMAQDAAMFGVAAATATATATLTFPEEAPESMTSAGG	180				
Db	121	LIATNLGQNTPAIAVNEAEYGEVMAQDAAMFGVAAATATATATLTFPEEAPLITNPGG	180				
Qy	181	LLEQAAVBEASDTAAANQLMNVFPQALQQLAQCTQGTTPSKLGLGWKTVSPHRSPLSN	240				
Db	181	LLEQAVAVEEADTAAANQLMNVFPQALQQLAQPTKSIWPFQDLSLWKAI:SPHLSPLSN	240				
Qy	241	MYSMANNHSMNTSGVSMYNTLTSLSMLKGFAPAAAAAQVQTAQNGVRAMSS----	296				
Db	241	IVSMLNNHSMNTSGVSMYASTLHSLKGFAP--AAQAQVETAQNGVQWSS:LGSQLGSSL	299				
Qy	297	GSGGLGGVGAANLGRAASVGSLSVQAAWAAANQAVTPAARALPLTSLTSAAREPGQWMLG	356				
Db	300	GSGGLGAGVGAANLGRAASVGSLSVQAAWAAANQAVTPAARALPLTSLTSAQAQTAPEHMLG	359				
Qy	357	GLPVGQMGARAG--GGLSGLRVPPRPVYMPHSPAPAG	391				
Db	360	GLPLGQGLTSGGSGFGGVSNALRMPPRAYVMPRPVPAAG					

RESULT 13
US-09-072-596-106
; Sequence 106, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Rasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 106:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 396 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 US-09-072-596-106

Query Match 84.8%; Score 1652.5; DB 4; Length 396;
 Best Local Similarity 84.9%; Pred. No. 1.4e-129;
 Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 DB 1 VVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 QY 61 SSAGLMVAASPPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 DB 61 SSAGLMVAASPPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 QY 121 LIATNLLGQNTPTAIVNEAEYGENWQAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
 DB 121 LIATNLLGQNTPTAIVNEAEYGENWQAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
 QY 181 LLEQAAVEEAADTAANQNMNVPQALQLOAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
 DB 181 LLEQAAVEEAADTAANQNMNVPQALQLOAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
 QY 241 MVSMMNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMS 296
 DB 241 IVSMNHNHVSMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMS 296
 QY 297 GSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLG 356
 DB 300 GSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLG 359
 QY 357 GLPVGMGARAG--GGLSGVLRVPPRYVMPHSPAAG 391
 DB 360 GLPLGQLTNSGGGPGGVSNALRMPPRAYVMPRPAAG 396

RESULT 14

US-09-072-967-111
 Sequence 111, Application US/09072967
 Patent No. 6592877

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, David C.
 APPLICANT: Campos-Neto, Antonio
 APPLICANT: Houghton, Raymond
 APPLICANT: Vedvick, Thomas S.
 APPLICANT: Twardzik, Daniel R.
 APPLICANT: Lodes, Michael J.
 APPLICANT: Hendrickson, Ronald C.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
 NUMBER OF SEQUENCES: 355
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMEUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/072,967
 FILING DATE: 05-MAY-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.411C9
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 111:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 396 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 US-09-072-967-111

Query Match 84.8%; Score 1652.5; DB 4; Length 396;
 Best Local Similarity 84.9%; Pred. No. 1.4e-129;
 Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
 QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 DB 1 VVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 QY 61 SSAGLMVAASPPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 DB 61 SSAGLMVAASPPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 QY 121 LIATNLLGQNTPTAIVNEAEYGENWQAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
 DB 121 LIATNLLGQNTPTAIVNEAEYGENWQAQDAAMFGYAAATATATATLTPFEAPEMTSAGG 180
 QY 181 LLEQAAVEEAADTAANQNMNVPQALQLOAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
 DB 181 LLEQAAVEEAADTAANQNMNVPQALQLOAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
 QY 241 MVSMMNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMS 296
 DB 241 IVSMNHNHVSMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMS 296
 QY 297 GSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLG 356
 DB 300 GSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLG 359
 QY 357 GLPVGMGARAG--GGLSGVLRVPPRYVMPHSPAAG 391
 DB 360 GLPLGQLTNSGGGPGGVSNALRMPPRAYVMPRPAAG 396

RESULT 15

US-08-818-112-109
 Sequence 109, Application US/08818112
 Patent No. 6290369

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, David C.
 APPLICANT: Campos-Neto, Antonio
 APPLICANT: Houghton, Raymond
 APPLICANT: Vedvick, Thomas S.
 APPLICANT: Twardzik, Daniel R.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
 NUMBER OF SEQUENCES: 153
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

Search completed: June 22, 2004, 17:27:32
Job time : 14.9576 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 36.2142 Seconds
(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-14
Perfect score: 1949
Sequence: 1 MVDGALPEINSARMYAGP.....SGVLRVPRPYVMPHSPRAG 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1949	100.0	391	12	US-09-886-349A-14
2	1949	100.0	391	14	US-10-193-002-102
3	1949	100.0	391	14	US-10-084-843-107
4	1949	100.0	391	14	US-10-098-732A-14
5	1949	100.0	596	9	US-09-287-849-26
6	1949	100.0	596	12	US-09-886-349A-20
7	1949	100.0	596	14	US-10-359-460-26
8	1949	100.0	596	14	US-10-098-732A-20
9	1949	100.0	600	9	US-09-287-849-22
10	1949	100.0	600	14	US-10-359-460-22
11	1949	100.0	723	15	US-10-369-983-2
12	1949	100.0	729	12	US-09-886-349A-18
13	1949	100.0	729	14	US-10-098-732A-18
14	1949	100.0	729	15	US-10-369-983-21
15	1949	100.0	729	15	US-10-369-983-22

16	1949	100.0	813	15	US-10-369-983-15	Sequence 15, Appl
17	1949	100.0	825	15	US-10-369-983-14	Sequence 14, Appl
18	1949	100.0	875	15	US-10-369-983-13	Sequence 13, Appl
19	1949	100.0	930	14	US-10-098-732A-65	Sequence 65, Appl
20	1949	100.0	930	15	US-10-369-983-12	Sequence 12, Appl
21	1949	100.0	1010	15	US-10-369-983-4	Sequence 4, Appl
22	1949	100.0	1016	15	US-10-369-983-18	Sequence 18, Appl
23	1949	100.0	1022	15	US-10-369-983-17	Sequence 17, Appl
24	1949	100.0	1154	15	US-10-369-983-16	Sequence 16, Appl
25	1944	99.7	729	9	US-09-287-849-2	Sequence 2, Appl
26	1944	99.7	729	12	US-09-886-349A-16	Sequence 16, Appl
27	1944	99.7	729	14	US-10-359-460-2	Sequence 2, Appl
28	1944	99.7	729	14	US-10-098-732A-16	Sequence 16, Appl
29	1944	99.7	729	15	US-10-359-459-2	Sequence 2, Appl
30	1934	99.2	391	12	US-09-872-186-8	Sequence 8, Appl
31	1652.5	84.8	396	14	US-10-193-002-106	Sequence 106, App
32	1652.5	84.8	396	14	US-10-084-843-111	Sequence 111, App
33	1583	81.2	393	12	US-10-282-122A-62455	Sequence 62455, A
34	1583	81.2	393	12	US-10-282-122A-64892	Sequence 64892, A
35	1486.5	76.3	359	14	US-10-193-002-104	Sequence 104, App
36	1486.5	76.3	359	14	US-10-084-843-109	Sequence 109, App
37	1187.5	60.9	358	9	US-09-287-849-8	Sequence 8, Appl
38	1187.5	60.9	358	14	US-10-359-460-8	Sequence 8, Appl
39	1187	60.9	263	12	US-09-886-349A-12	Sequence 12, Appl
40	1187	60.9	263	14	US-10-193-002-92	Sequence 92, Appl
41	1187	60.9	263	14	US-10-084-843-91	Sequence 91, Appl
42	1187	60.9	263	14	US-10-098-732A-12	Sequence 12, Appl
43	766.5	39.3	400	9	US-09-073-009-126	Sequence 126, App
44	766.5	39.3	400	9	US-09-793-306-126	Sequence 126, App
45	742.5	38.1	405	12	US-10-282-122A-62027	Sequence 62027, A

ALIGNMENTS

RESULT 1

US-09-886-349A-14
; Sequence 14, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886.349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB39 (TbH9FL)
US-09-886-349A-14

Query Match 100.0%; Score 1949; DB 12; Length 391;
Best Local Similarity 100.0%; Pred. No. 1e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVDGALPEINSARMYAGPGASLVAAQWDSVADLFSAAAFQSVVGLTVGSMIG 60
Db	1	MVDGALPEINSARMYAGPGASLVAAQWDSVADLFSAAAFQSVVGLTVGSMIG 60
Qy	61	SSAGLWVAASPYVAMSVTAGQBELTAAQVRVAAAAAYETAYGLTVPVPPVIAENRAELMI 120
Db	61	SSAGLWVAASPYVAMSVTAGQBELTAAQVRVAAAAAYETAYGLTVPVPPVIAENRAELMI 120

QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLPPPEAPMTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQNMNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
DB 181 LLEQAAVEEASDTAAANQNMNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
QY 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSLGSSG 300
DB 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSLGSSG 300
QY 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
DB 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
QY 361 GQMGARAGGGLSGVLVPPRPYPVPHSPAAG 391
DB 361 GQMGARAGGGLSGVLVPPRPYPVPHSPAAG 391

RESULT 2

US-10-193-002-102

; Sequence 102, Application US/10193002

; Publication No. US20030135026A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; Skeiky, Yasir A.W.

; Dillon, Davin C.

; Campos-Neto, Antonia

; Houghton, Raymond

; Vedvick, Thomas S.

; Twardzik, Daniel R.

; Lodes, Michael J.

; Hendrickson, Ronald C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS

; NUMBER OF SEQUENCES: 350

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/193,002

; FILING DATE: 10-Jul-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/072,596

; FILING DATE: 05-MAY-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.417C9

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 102:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 391 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 102:

; US-10-193-002-102

Query Match 100.0%; Score 1949; DB 14; Length 391;
Best Local Similarity 100.0%; Pred. No. 1e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYDFGALPPEINSAARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVWGLTVGSGWIG 60
DB 1 MYDFGALPPEINSAARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVWGLTVGSGWIG 60

QY 61 SSAGLWVAAAAPFYVAMSVTAGQAELETAQVRAVAAAAYETAYGLTVPPVTAENRAELMI 120
DB 61 SSAGLWVAAAAPFYVAMSVTAGQAELETAQVRAVAAAAYETAYGLTVPPVTAENRAELMI 120

QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180

QY 181 LLEQAAVEEASDTAAANQNMNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
DB 181 LLEQAAVEEASDTAAANQNMNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240

QY 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSLGSSG 300
DB 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAPAAAQAVQTAQNGVRAMSSLGSSG 300

QY 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
DB 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360

QY 361 GQMGARAGGGLSGVLVPPRPYPVPHSPAAG 391
DB 361 GQMGARAGGGLSGVLVPPRPYPVPHSPAAG 391

RESULT 3

US-10-084-843-107

; Sequence 107, Application US/10084843

; Publication No. US20030143243A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; Skeiky, Yasir A.W.

; Dillon, Davin C.

; Campos-Neto, Antonio

; Houghton, Raymond

; Vedvick, Thomas S.

; Twardzik, Daniel R.

; Lodes, Michael J.

; Hendrickson, Ronald C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS

; NUMBER OF SEQUENCES: 355

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/084,843

; FILING DATE: 25-Feb-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/072,967

; FILING DATE: 05-MAY-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

Query Match	Score 1949;	DB 14;	Length 391;
100.0%			

Db 9 MVDFGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 68
QY 61 SSAGLMVAASPYVWMSVTAQAGBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 69 SSAGLMVAASPYVWMSVTAQAGBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
QY 121 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
Db 129 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 188
QY 181 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
Db 189 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 248
QY 241 MVSANNHMTNSGVSMNTLSMLKGFAPAPAAAQAVCTAAQNGVRAMSSLSGSSG 300
Db 249 MVSANNHMTNSGVSMNTLSMLKGFAPAPAAAQAVCTAAQNGVRAMSSLSGSSG 308
QY 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTTPAARALPLTSLTSAERPGQMGLPLV 360
Db 309 LGGGVAANLGRAASVGSLSVPQAWAANQAVTTPAARALPLTSLTSAERPGQMGLPLV 368
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 369 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 399

RESULT 6

US-09-886-349a-20
; Sequence 20, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-886-349a-20

Query Match 100.0%; Score 1949; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.8e-145; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDFGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Db 9 MVDFGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 68
QY 61 SSAGLMVAASPYVWMSVTAQAGBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 69 SSAGLMVAASPYVWMSVTAQAGBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
QY 121 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
Db 129 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
QY 181 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240

Db 189 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 248
QY 241 MVSANNHMTNSGVSMNTLSMLKGFAPAPAAAQAVCTAAQNGVRAMSSLSGSSG 300
Db 249 MVSANNHMTNSGVSMNTLSMLKGFAPAPAAAQAVCTAAQNGVRAMSSLSGSSG 308
QY 301 LGGGVAANLGRAASVGSLSVPQAWAANQAVTTPAARALPLTSLTSAERPGQMGLPLV 360
Db 309 LGGGVAANLGRAASVGSLSVPQAWAANQAVTTPAARALPLTSLTSAERPGQMGLPLV 368
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 369 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 399

RESULT 7

US-10-359-460-26
; Sequence 26, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-10-359-460-26

Query Match 100.0%; Score 1949; DB 14; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.8e-145; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDFGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Db 9 MVDFGALPPEINSARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 68
QY 61 SSAGLMVAASPYVWMSVTAQAGBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 69 SSAGLMVAASPYVWMSVTAQAGBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
QY 121 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
Db 129 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
QY 181 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
Db 189 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 248

QY 241 MVSMAHNMHMTNSGVSMNTNTLSMLKGFAPAAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 300
DB 249 MVSMAHNMHMTNSGVSMNTNTLSMLKGFAPAAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 308
QY 301 LGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
DB 309 LGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 368

QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
DB 369 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 399

RESULT 8

US-10-098-732A-20
; Sequence 20, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TBH9-Ra35 (designated MTB59F)
US-10-098-732A-20

Query Match 100.0%; Score 1949; DB 14; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.8e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68

QY 61 SSAGLMVAASAPYVAMWSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
DB 69 SSAGLMVAASAPYVAMWSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 128

QY 121 LIATNLLGQNTPTAIAVNEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
DB 129 LIATNLLGQNTPTAIAVNEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188

QY 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQAQPTQGTTPSSKLGGLWKTVSPHRSPTSN 240
DB 189 LLEQAAAVEEASDTAAANQLMNNVPOALQQAQPTQGTTPSSKLGGLWKTVSPHRSPTSN 248

QY 241 MVSMAHNMHMTNSGVSMNTNTLSMLKGFAPAAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 300
DB 249 MVSMAHNMHMTNSGVSMNTNTLSMLKGFAPAAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 308

QY 301 LGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
DB 309 LGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 368

QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
DB 369 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 399

RESULT 9
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

Query Match 100.0%; Score 1949; DB 9; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.8e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
QY 61 SSAGLMVAASAPYVAMWSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
DB 69 SSAGLMVAASAPYVAMWSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 128
QY 121 LIATNLLGQNTPTAIAVNEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
DB 129 LIATNLLGQNTPTAIAVNEAEYGENWQAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
QY 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQAQPTQGTTPSSKLGGLWKTVSPHRSPTSN 240
DB 189 LLEQAAAVEEASDTAAANQLMNNVPOALQQAQPTQGTTPSSKLGGLWKTVSPHRSPTSN 248
QY 241 MVSMAHNMHMTNSGVSMNTNTLSMLKGFAPAAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 300
DB 249 MVSMAHNMHMTNSGVSMNTNTLSMLKGFAPAAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 308
QY 301 LGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 360
DB 309 LGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALPLTSLTSAABRGPGQMLGGLPV 368
QY 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
DB 369 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 399

RESULT 10
US-10-359-460-22
; Sequence 22, Application US/10359460


```
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72FmutSA
; OTHER INFORMATION: (Ra12-TbHp-Ra35MutSA)
US-09-886-349a-18

Query Match      100.0%; Score 1949; DB 12; Length 729;
Best Local Similarity 100.0%; Pred. No. 2.3e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Db 142 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 201
QY 61 SSAGLMVAAAAPYVAMVSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPPVIAENRAELMI 120
Db 202 SSAGLMVAAAAPYVAMVSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPPVIAENRAELMI 261
QY 121 LIATNLLGQNTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 180
Db 262 LIATNLLGQNTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 321
QY 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
Db 322 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 381
QY 241 MYSVANNHMSMTNSGVSMNTLSSMLKGFAPAPAAAQAQVTPAARALPLTSLTSAERGFQMLGGLPV 300
Db 382 MYSVANNHMSMTNSGVSMNTLSSMLKGFAPAPAAAQAQVTPAARALPLTSLTSAERGFQMLGGLPV 441
QY 301 LGGGVAANLGRAASVGSLSVPOQAWAANAQAVTPAARALPLTSLTSAERGFQMLGGLPV 360
Db 442 LGGGVAANLGRAASVGSLSVPOQAWAANAQAVTPAARALPLTSLTSAERGFQMLGGLPV 501
QY 361 GQMGARAGGGLSGVLRVPPRPVYMPHSPAAG 391
Db 502 GQMGARAGGGLSGVLRVPPRPVYMPHSPAAG 532

RESULT 13
US-10-098-732A-18
; Sequence 18, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderman, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098, 732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72FmutSA
US-10-098-732A-18

Query Match      100.0%; Score 1949; DB 14; Length 729;
Best Local Similarity 100.0%; Pred. No. 2.3e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Db 142 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 201
QY 61 SSAGLMVAAAAPYVAMVSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPPVIAENRAELMI 120
Db 202 SSAGLMVAAAAPYVAMVSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPPVIAENRAELMI 261
QY 121 LIATNLLGQNTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 180
Db 262 LIATNLLGQNTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 321
QY 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
Db 322 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 381
QY 241 MYSVANNHMSMTNSGVSMNTLSSMLKGFAPAPAAAQAQVTPAARALPLTSLTSAERGFQMLGGLPV 300
Db 382 MYSVANNHMSMTNSGVSMNTLSSMLKGFAPAPAAAQAQVTPAARALPLTSLTSAERGFQMLGGLPV 441
QY 301 LGGGVAANLGRAASVGSLSVPOQAWAANAQAVTPAARALPLTSLTSAERGFQMLGGLPV 360
Db 442 LGGGVAANLGRAASVGSLSVPOQAWAANAQAVTPAARALPLTSLTSAERGFQMLGGLPV 501
QY 361 GQMGARAGGGLSGVLRVPPRPVYMPHSPAAG 391
Db 502 GQMGARAGGGLSGVLRVPPRPVYMPHSPAAG 532

RESULT 14
US-10-369-983-21
; Sequence 21, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderman, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72F
US-10-369-983-21

Query Match      100.0%; Score 1949; DB 15; Length 729;
Best Local Similarity 100.0%; Pred. No. 2.3e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Db 142 MVDGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 201
QY 61 SSAGLMVAAAAPYVAMVSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPPVIAENRAELMI 120
Db 202 SSAGLMVAAAAPYVAMVSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPPVIAENRAELMI 261
QY 121 LIATNLLGQNTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 180
Db 262 LIATNLLGQNTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 321
QY 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
Db 322 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 381
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:59:04 ; Search time 10.9397 Seconds
(without alignments)
3438.018 Million cell updates/sec

Title: US-09-886-349A-14
Perfect score: 1949
Sequence: 1 MVDFGALPPEINSARMYAGP.....SCVLVRPPRYVMPHSPAAG 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1949	100.0	391	2 B70608	probable PPE prote
2	1656.5	85.0	386	2 H70741	probable PPE prote
3	1583	81.2	393	2 C70568	probable PPE prote
4	779.5	40.0	393	2 G70929	probable PPE prote
5	751	38.5	409	2 A70932	probable PPE prote
6	737.5	37.8	403	2 H70931	probable PPE prote
7	736	37.8	423	2 B70931	probable PPE prote
8	731	37.5	421	2 H87056	PPE-family protein
9	705	36.2	408	2 G70925	probable PPE prote
10	702	36.0	391	2 B70625	probable PPE prote
11	702	36.0	413	2 F70560	probable PPE prote
12	688.5	35.3	463	2 C70931	probable PPE prote
13	682.5	35.0	468	2 B70932	probable PPE prote
14	675	34.6	380	2 A70646	probable PPE prote
15	668.5	34.3	394	2 G70881	probable PPE prote
16	667.5	34.2	385	2 H70903	probable PPE prote
17	635	32.6	350	2 H70929	probable PPE prote
18	633.5	32.5	365	2 E70929	probable PPE prote
19	615.5	31.6	402	2 C70882	probable PPE prote
20	603	30.9	423	2 C70582	probable PPE prote
21	597.5	30.7	391	2 D70922	probable PPE prote
22	589	30.2	406	2 E70675	probable PPE prote
23	588	30.2	394	2 A70504	probable PPE prote
24	575.5	29.5	391	2 D70663	probable PPE prote
25	495.5	25.4	3300	2 D70575	probable PPE prote
26	473	24.3	180	2 G70834	probable PPE prote
27	452.5	23.2	3716	2 E70969	probable PPE prote
28	448.5	23.0	580	2 G70570	probable PPE prote
29	444	22.8	346	2 H70874	probable PPE prote

30	443	22.7	963	2 B70524	probable PPE prote
31	441	22.6	487	2 C70830	probable PPE prote
32	439.5	22.6	655	2 A70931	probable PPE prote
33	439.5	22.0	479	2 D70676	probable PPE prote
34	438.5	22.0	1053	2 B70987	probable PPE prote
35	436.5	21.9	443	2 C70780	probable PPE prote
36	433.5	21.7	2523	2 F70846	probable PPE prote
37	422.5	21.7	615	2 E70663	probable PPE prote
38	418	21.4	678	2 A70762	probable PPE prote
39	418	21.4	3157	2 B70969	probable PPE prote
40	415.5	21.3	618	2 F70552	probable PPE prote
41	415	21.3	582	2 H70675	probable PPE prote
42	414.5	21.3	987	2 E70808	probable PPE prote
43	412	21.1	590	2 E70946	probable PPE prote
44	412	21.1	645	2 F70825	probable PPE prote
45	409.5	21.0	409	2 A70647	probable PPE prote

ALIGNMENTS

RESULT 1

B70608
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70608
R:Coile, S.I.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Genclies, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whithead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: B70608
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-391 <COL>
A:Cross-references: GB:Z93777; GB:AL123456; NID:g3261726; PIDN:CAB07839.1; PID:e311073; I
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match		100.0%;	Score 1949;	DB 2;	Length 391;
Best Local Similarity		100.0%;	Pred. No. 3.5e-108;		
Matches 391;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MVDFGALPPEINSARMYAGPGSASIVAAQWDSVADLFSASAFQSVWGLTVGSWIG	60		
DB	1	MVDFGALPPEINSARMYAGPGSASIVAAQWDSVADLFSASAFQSVWGLTVGSWIG	60		
QY	61	SSAGLWVAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI	120		
DB	61	SSAGLWVAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI	120		
QY	121	LIATNLGQNTPATVNEAEYGEWQAQAAAFGVAATATATATLLPFEAPENTSAGG	180		
DB	121	LIATNLGQNTPATVNEAEYGEWQAQAAAFGVAATATATATLLPFEAPENTSAGG	180		
QY	181	LLEQAAVAEEASDTAAANQLMNNVPOALQQAQTOGTTPSKLGLWKTVPSPHRSPISN	240		
DB	181	LLEQAAVAEEASDTAAANQLMNNVPOALQQAQTOGTTPSKLGLWKTVPSPHRSPISN	240		
QY	241	MVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQCTAAQNGVRAMSSLGSLGSSG	300		
DB	241	MVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQCTAAQNGVRAMSSLGSLGSSG	300		
QY	301	LGCGVAANLGRAASVGSLSVQAAANQAVTPAARALPLTSLTSAAPRGQMLGGLPV	360		
DB	301	LGCGVAANLGRAASVGSLSVQAAANQAVTPAARALPLTSLTSAAPRGQMLGGLPV	360		
QY	361	GOMGARAGGSLSGVLRVPPRYVMPHSPAAG	391		

```
Db 361 GQMGARAGGGLGVLRVPPRPVMPHSPAAG 391
RESULT 2
H70741
A:Molecule type: DNA
A:Residues: 1-393 <COL>
A:Cross-references: GB:2755390; GB:AL123456; NID:g3261766; PIDN:CAB08702.1; PID:e316074; I
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 81.2%; Score 1583; DB 2; Length 393;
Best Local Similarity 81.5%; Pred. No. 1.5e-86;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGWSWIG 60
DB 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGWSWIG 60
QY 61 SSAGLWMAAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLWMAAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGONTPTAVNAEYGEWMAQDAAMFGYAAATATATATATLTPFEAPEMTSAGG 180
DB 121 LIATNLLGONTPTAVNAEYGEWMAQDAAMFGYAAATATATATATLTPFEAPEMTSAGG 180
QY 181 LLEQAAAVEASDTAAANQLMNNVPOALQQAQPTQGTTPSSKLGGLWKTVPSPHSPI 240
DB 181 LLEQAAAVEASDTAAANQLMNNVPOALQQAQPTQGTTPSSKLGGLWKTVPSPHSPI 240
QY 241 MVSMAHNSMTNSGVSMNTLSMLKGFAPAAAQAVQTAQNGVRAMSS----LGSSL 296
DB 241 MVSMAHNSMTNSGVSMNTLSMLKGFAPAAAQAVQTAQNGVRAMSS----LGSSL 296
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERPGQMLG 356
DB 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERPGQMLG 356
QY 357 GLPVQMGARAG--GGLSGVLRVPPRPVMPHSPAAG 391
DB 357 GLPVQMGARAG--GGLSGVLRVPPRPVMPHSPAAG 391
QY 360 GLPLGH-SVNAAGSGINNLRVPAAYALPTPTAAG 393
DB 360 GLPLGH-SVNAAGSGINNLRVPAAYALPTPTAAG 393

RESULT 4
G70929
Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70929
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70929
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-393 <COL>
A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAA17711.1; PID:e125460C
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 40.0%; Score 779.5; DB 2; Length 393;
Best Local Similarity 43.7%; Pred. No. 4.3e-39;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGWSWIG 61
DB 2 VDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGWSWIG 61
QY 61 SSAGLWMAAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 121
DB 61 SSAGLWMAAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 121
QY 121 LIATNLLGONTPTAVNAEYGEWMAQDAAMFGYAAATATATATATLTPFEAPEMTSAGG 180
DB 121 LIATNLLGONTPTAVNAEYGEWMAQDAAMFGYAAATATATATATLTPFEAPEMTSAGG 180
QY 181 LLEQAAAVEASDTAAANQLMNNVPOALQQAQPTQGTTPSSKLGGLWKTVPSPHSPI 240
DB 181 LLEQAAAVEASDTAAANQLMNNVPOALQQAQPTQGTTPSSKLGGLWKTVPSPHSPI 240
QY 241 MVSMAHNSMTNSGVSMNTLSMLKGFAPAAAQAVQTAQNGVRAMSS----LGSSL 296
DB 241 MVSMAHNSMTNSGVSMNTLSMLKGFAPAAAQAVQTAQNGVRAMSS----LGSSL 296
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERPGQMLG 356
DB 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERPGQMLG 356
QY 357 GLPVQMGARAG--GGLSGVLRVPPRPVMPHSPAAG 391
DB 357 GLPVQMGARAG--GGLSGVLRVPPRPVMPHSPAAG 391
QY 360 GLPLGH-SVNAAGSGINNLRVPAAYALPTPTAAG 393
DB 360 GLPLGH-SVNAAGSGINNLRVPAAYALPTPTAAG 393

RESULT 3
C70568
Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: C70568
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: C70568
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
```

QY 122 IATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAEMTSAGGL 181
Db 121 ISTNVFGQNTSALTAABAEAYGEWMAQDSAAVAYAGSSASASA-VTFEFTPTQIANPTAQ 179
QY 182 LEQAAAVEBASDTAAA--NOLMNNVPQALQOLAQPTGCTTPSPKGLGINK----- 229
Db 180 GTQAAAVATAAGTAQTLTTEMTIGLNLQSLTSPILQSS--NGPLSMLWQLLFGTNPFFT 238
QY 230 -----TVSPHRSPTSNMVMNMMNMTNSGVSMNTLSSMLKGFAPAAAQAQVATAA 282
Db 239 SISALLTDLCQPYASFFYNTEGLPYFISIGMGNFIQSAKTL-GLIGSAAFAAVA-----AA 292
QY 283 QNGVRAMWSLGSLSGSGGLGGVAAVNLGRAASVGSLSVQAAMAA--ANQAVTPAARALPLT 341
Db 293 GDAAKGLPGLGMLG---GGPVAAGLGNAAAYGKLSVPPVMSGPLPGSVTPCAAPLPVS 348
QY 342 SLTSAAEERGQOMLGLPVQCMGARAGGSLGVLRPVPPYVMPHPSPAAG 391
Db 349 TVSAAPAEAPGSLGGLPL-----AGAGGAGAGP-RYGFRPTVMARPPFAG 393

RESULT 5
A70932
Probable pPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70932
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Cole, S.T.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70932
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-409 <COL>
A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAA17729.1; PID:e125461
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 38.5%; Score 751; DB 2; Length 409;
Best Local Similarity 44.7%; Pred. No. 2.2e-37;
Matches 177; Conservative 53; Mismatches 138; Indels 28; Gaps 10;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAAQWDSVADLPSAASAFQSVWGLTVGWSIGS 61
Db 1 MDFGALPPEINSGRMYAGPGSGPLAAAADWALAAELYGAAAYSGSTIEGLTVAPWMPG 60
QY 62 SAGLWMAAASPYVAWMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVPVIAENRAELMIL 121
Db 61 SSITMAAAVAPYVAWISVTAGAEQACQAKIAGVYETAFATVPPVPIEANKALLMSL 120
QY 122 IATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEWTSAGGL 181
Db 121 VATNIFGONTPAIAATEAHYAEMWMAQDAAMFGYAGSSATA--SQLAPFSEPPPTNPST 179
QY 182 LEQAAAVEBASDTAAA-----NOLMNNVPQALQOLAQPTGCTTPSPKGLGWLKTVSPH 234
Db 180 AAQSAVVAQAAGAAASDDITQLSLSLPTLQSLA--TTATATSASAG--WDIV--- 232
QY 235 RSPISNNVMNMMNMTNSGVSMNTLSSMLKGFAPAAAQAQVATAAQNQVGRAM 289
Db 233 LQSTITLANLTGPYSIIGLCAIPGGWMLTFGQILGLAQNAPGVAALLGPKAAAGALSPL 292
QY 290 SSL--GSSLGS--SGLGGVAAVNLGRAASVGSLSVQAAMAAANQAVTPAARALPLTSLTSA- 346
Db 293 APLSGGVIGDITPLGGGATGGIARAIYVGSLSVPPQGWAEAPVWRAVASVLPGTGAAPAL 352
QY 347 AERPGQWLGLPVQCMGARAGGGL-----SGVLRV 377

[illegible]

Db 180 AGQAAATVSTVPLATTAAPVQLQLQSGSTSLIPWYSALQWLAENLLGLTPDNRMTIV 239
QY 225 -----GGLWKTVPSPHSPISSNVSMANNHMTNSGVSMNTLSSMLKGFAPAAAA 275
Db 240 RLIGISYFDEGL-----LQFASLAQAQIPGTGAG--DSGSSVLDSWGTIFA 287
QY 276 QAVQTAQNGVRAMSSL--GSSLSG-----SGLGGVAAANLGRAASVGSLS 319
Db 288 -----GPRASPSPVAGGAVGGVQTQPPQYWMALDRSIGSVSAALGKGSAGSLS 338
QY 320 VPCWAAANQAVTPAARALP--LTSLSAERPGQMLGGLPVGQMGARAGGSLGVLR 376
Db 339 VPPDMAARAWANPAARLPGDDVTALRGTAENA---LLRGFPMSAGOSTGGF--VHK 393
QY 377 VPPRPVYMPHSPAAG 391
Db 394 YGFLAVMQRPFPAG 408

RESULT 10
B70625
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70625
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: B70625
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-391 <COL>
A: Cross-references: GB:292539; GB:AL123456; NID:G3261714; PIDN: CAB06873.1; PID: e304546;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: PPE

Query Match 36.0%; Score 702; DB 2; Length 391;
Best Local Similarity 42.3%; Pred. No. 1.6e-34;
Matches 166; Conservative 62; Mismatches 150; Indels 14; Gaps 7;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIGS 61
Db 1 MDFGALPPEINSARMYAGAGAGPMMAAGAAWNGLAALGELGTTAASYSVITRLTTESWMP 60
QY 62 SAGLWAAASPYVAVWVSTAGQAEELTAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 121
Db 61 ASWAVAAQPYLAWITYTAAAHAGSQMASAAAYEAAYAVTVPFVVAANRALLAAL 120
QY 122 IATNLLGONTPTAIVNAEYGEVMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGL 181
Db 121 VATNLGINTPATMATEALYAEWMAQDALAMYAAASG--AAGMLQLSPSPQTTNPGL 179
QY 182 LEQAAVEASDPAARANQ-----LMNVQALQQLAQPTQGTTPSSKLGKWKTVSPHRS 236
Db 180 AQSAIVGSAATAAVNQVNSVDLSSLPNAVSGLASPVTSVLDSTGSGIITADIALIA 239
QY 237 P--ISNVSMANNHMS--MTNSGVSMNTLSSMLKGFAPAAAAQVQTAQNGVRAMSSLSG 293
Db 240 TTPVANIINSVNTAAWYNAIPTAIFLANALNSGAPVAIEGALEAAEG--AASAAA 296
QY 294 SSLGSSGLGGVAAANI GRAASVGSLSVPOQWAAANQAVTPAARALPLTSLTSAERGPQ 353
Db 297 AGLADSVTPAGLGSIGLEATLVGLRSPAAWSTAAPATTAGATALEGSGWTVAEEA-QF 355
QY 354 MLGGLPVQMGARAGGSLGVLRVPPRPVMP 385
Db 356 VTGWMP-GHSAAXKGIGAGVPRYGRKPTMP 386

RESULT 11
F70560
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: F70560
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: F70560
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-413 <COL>
A: Cross-references: GB:295436; GB:AL123456; NID:G3261770; PIDN: CAB08826.1; PID: e316565;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: PPE

Query Match 36.0%; Score 702; DB 2; Length 413;
Best Local Similarity 39.1%; Pred. No. 1.7e-34;
Matches 168; Conservative 61; Mismatches 145; Indels 56; Gaps 7;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
Db 1 MLDFAQLPPEVNSALMYAGPGSGPMLAAAAAWEALAELOTTASTYDALITGLADGPWQ 60
QY 61 SSAGLWAAASPYVAVWVSTAGQAEELTAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
Db 61 SSAASMVAAATPQVAVLNRSTAGQAEQAGSQAVAAASAYEAAFFATVPPPEIAANRALLMA 120
QY 121 LIATNLLGONTPTAIVNAEYGEVMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
Db 121 LLATNLLGONTPTAIVNAEYGEVMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 179
QY 181 LLEQAAVEASDPAARANQVQALQQLAQPTQGTTPSSKLGKWKTVSPHRSPTSN 240
Db 180 LASQAAASVQAVSGAANAQALTDIPKAL-----FGLSGIFTNPPMLTDLGK 226
QY 241 MVSMAHHMTNSGVSMNTLSSMLKGFAPAA---AAQAVQT----- 280
Db 227 ALGLTGWSSDGGSLIVGGVGLDFVQGVTSQSLDASVAMDTFGKVVSPARLMVTQPKD 286
QY 281 -----AAQNGVRAMSSLSGSSLSGSGVAAANLGRAASVGSLSVPOA 323
Db 287 YFGLAHDLPKWASEGAKAAGEAAKALPAAVPAIPISAGL-SCVAGAVGQAASVGLKVPVAV 345
QY 324 WAAANQAVTPAARALPLTSLTSAERGPQMLGGLPVQMGARAGGSLGVLR--RVPPRP 381
Db 346 WTATTPAASPVAALAAASNLGAAAAAEGSTHAFGGMPL--MGSGAGRAFNNFAAPRYGPKP 403
QY 382 YVPHSPSPAAG 391
Db 404 TVIAQPPPAGG 413

RESULT 12
C70931
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: C70931
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: C70932

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-463 <COL>

A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAAL1773.1; PID:e125462

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match 35.3%; Score 688.5; DB 2; Length 463;

Best Local Similarity 42.8%; Pred. No. 1.3e-33;

Matches 174; Conservative 48; Mismatches 126; Indels 49; Gaps 11;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVWGLTVGWSIGS 61

DB 1 MDFGVLPPEINSGRMYAGPGSGPMLAAAAWDGLATELQSTAADYGSVIVSLT-GVMSGQ 59

QY 62 SAGLMVAAAAPYVAVMSVTAQAEELTAQVRAAAAYETAVGLTVPPVIAENRAELMIL 121

DB 60 SSGTMAAAAPYVAVMSVTAQAEELTAQVRAAAAYETAVGLTVPPVIAENRAELMIL 119

QY 122 IATNLLGQNTPAIAVNEAEYEMMAQDAAMFGYAAATATATATATLPPFEAPEMTSAGGL 181

DB 120 AATNIFGQNTGAIAAAEARYAEMMAQDAAMFGYAAATATATATATLPPFEAPEMTSAGGL 178

QY 182 LEQAAVVEASDTAAANQLMNNVPOALQLOAQPTGCTTPSSKLGGLWKTYS--PHRSP1- 238

DB 179 ATQGVAVAAQVAGSAGN-ARSLVSEVLEFLA--TAGTNNKTVASLMNAVTVGPVYASSVY 235

QY 239 -----SNMYSMANNHMTNSGVSMNTNTLSSMLKGFAPAAAQAVQTAQAQNGVRA 288

DB 236 NSMLGLGFAESKMWLPANDIVISIFGWFQFQFNFVTFNPDLPK----- 283

QY 289 MSSIGSSLG-----SSGLGG---GVAANLGRAASVGLSVPOQWAAANQAVTPAARALPL 340

DB 284 -SALGAGLGLRSLSSGLGSTAPASAGASCAGSVGMSVPPSWAAATPAIRTVAAFVSS 342

QY 341 TSLTFS--AAERPGQML-----GGLPVGMGARGGGLSGVLRV 377

DB 343 TGLQVPAAPAISEGLSSQWALASVAGGALGGAAARATGGLGGRV 389

RESULT 13

B70932

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: B70932

R:Colle, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-468 <COL>

A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAAL1773.1; PID:e125462

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match 35.0%; Score 682.5; DB 2; Length 468;

Best Local Similarity 40.9%; Pred. No. 2.9e-33;

Matches 164; Conservative 59; Mismatches 125; Indels 53; Gaps 9;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVWGLTVGWSIGS 61

DB 1 MDFGLQPPETISGMYLPGAGPMLAAAVWDGLAELQSMASASIASIVEGMASESWLGP 60

Search completed: June 22, 2004, 17:24:52
Job time : 11.9397 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:46:13 ; Search time 6.50724 Seconds
(without alignments)
3128.737 Million cell updates/sec

Title: US-09-886-349A-14
Perfect score: 1949
Sequence: 1 MVDGALPPEINSARMYAGP.....SGVLRVPPRYVMHPSPAAG 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1656.5	85.0	396	1 YD61_MYCTU	Q11031 mycobacteri
2	705	36.2	408	1 YS92_MYCTU	Q10813 mycobacteri
3	688.5	35.3	483	1 Y102_MYCTU	O53951 mycobacteri
4	444	22.8	487	1 Y442_MYCTU	P42811 mycobacteri
5	426.5	21.9	443	1 Y878_MYCTU	Q10540 mycobacteri
6	418	21.4	678	1 YF48_MYCTU	Q10778 mycobacteri
7	364.5	18.7	408	1 SRA_MYCLE	Q07297 mycobacteri
8	324.5	16.6	463	1 Y096_MYCTU	Q10892 mycobacteri
9	324	16.6	434	1 YU18_MYCTU	P31500 mycobacteri
10	321.5	16.5	435	1 YU21_MYCTU	O53268 mycobacteri
11	233.5	12.0	178	1 Y125_MYCTU	O06246 mycobacteri
12	217.5	11.2	176	1 Y125_MYCTU	Q50702 mycobacteri
13	153.5	7.9	860	1 ELS_MOUSE	P54320 mus musculu
14	151.5	7.8	232	1 YV26_MYCTU	Q99372 rattus norv
15	150	7.7	864	1 ELS_RAT	P15502 homo sapien
16	143.5	7.4	730	1 ELS_HUMAN	P47033 saccharomyc
17	139.5	7.2	881	1 PRY3_YEAST	O60641 homo sapien
18	135	6.9	907	1 A180_HUMAN	P28284 herpes simp
19	133.5	6.8	825	1 ICP0_HSV2H	P18305 chilo iride
20	133.5	6.8	2432	1 Y10R_IRV6	P76072 escherichia
21	132	6.8	1120	1 STFR_ECOLI	P35658 homo sapien
22	132	6.8	2090	1 N214_HUMAN	Q05233 mycobacteri
23	131	6.7	836	1 VQ26_BPML5	P12021 sus scrofa
24	131	6.7	1150	1 APMU_PIG	P52172 chlamydomon
25	129	6.6	1783	1 RAA3_CHLRE	Q24856 nototheria
26	128	6.6	779	1 SRP_CHROME	Q10169 chitosacch
27	127.5	6.5	790	1 ANP_NOTCO	Q24523 drosophila
28	126	6.5	354	1 YAU8_SCHPO	P33240 homo sapien
29	125.5	6.4	1211	1 BUN2_DROME	P51611 mesocricetu
30	125	6.4	577	1 GST2_HUMAN	Q50597 mycobacteri
31	124	6.4	2090	1 HFC1_MESAU	P71789 mycobacteri
32	123.5	6.3	677	1 Y136_MYCTU	
33	123	6.3	432	1 YF10_MYCTU	

34	123	6.3	1140	1 YM96_YEAST	Q04893 saccharomyc
35	122	6.3	394	1 HYF1_ALCEU	P45805 alcaligenes
36	121.5	6.2	1025	1 SLAP_CAUCR	P35828 caulobacter
37	121.5	6.2	1845	1 Z236_HUMAN	Q9L36 homo sapien
38	121	6.2	1199	1 P121_RAT	P32591 rattus norv
39	119.5	6.1	635	1 HMLA_DROME	P10105 drosophila
40	119	6.1	915	1 A180_RAT	Q05140 rattus norv
41	118.5	6.1	580	1 EXPR_XANCP	P23314 xanthomonas
42	118.5	6.1	2038	1 FSH_DROME	P13709 drosophila
43	118	6.1	1508	1 BCSC_XANAC	P88938 xanthomonas
44	117	6.0	444	1 Y808_CHLPN	Q92798 chlamydia p
45	117	6.0	652	1 PICA_HUMAN	Q13492 homo sapien

ALIGNMENTS

RESULT 1	YD61_MYCTU	STANDARD:	PRT:	396 AA.
ID	YD61_MYCTU			
AC	Q11031;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Hypothetical PPE-family protein RV1361C/MT1406.			
GN	RV1361C OR MT1406 OR MYC02B10.25C.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RX	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Biglieri K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornsby T., Jagels K., Krogn A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RX	MEDLINE=22206494; PubMed=12218036;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey B.,			
RA	Kolodny J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;			
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	J. Bacteriol. 184:5479-5490(2002).			
CC	!- SIMILARITY: Belongs to the mycobacterial PPE family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z75555; CAA99966.1; -.			
DR	EMBL; ABC07013; AAK45669.1; -.			
DR	PIR; H70741; H70741.			
DR	TIGR; MT1406; -.			
DR	Tuberculist; RV1361C; -.			
DR	InterPro; IPR000030; Microbac_PPE.			

```

DR Pfam: PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 158 159 TA -> AT (IN REF. 2).
SQ SEQUENCE 396 AA; 40015 MW; 6AFAE0D7B5F668D0 CRC64;

Query Match      85.0%; Score 1656.5; DB 1; Length 396;
Best Local Similarity 85.1%; Pred. No. 1.3e-89;
Matches 338; Conservative 18; Mismatches 34; Indels 7; Gaps 3;

QY 1 MVDGALPPEINSMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
DB 1 MVDGALPPEINSMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY 61 SSAGLMVAASPYVAMSVTAQAGELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAASPYVAMSVTAQAGELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFCAATATATATLTPFEAPEMTSAGG 180
DB 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFCAATATATLTPFEAPEMTSAGG 180
QY 181 LLEQAAVEASDTAAANQLMNVPQALQQAQTOGTPTSSKLGGLWKTVPVHRSPISN 240
DB 181 LLEQAAVEASDTAAANQLMNVPQALQQAQTOGTPTSSKLGGLWKTVPVHRSPISN 240
QY 241 MVSMAHNMSTNGSVMTNTLSMLKGFAPAAAAQVTAQAQNGVEAMSS----LQSSSL 296
DB 241 IVSMLNHVSTNGSVMSATLHMLKGFAP-AAAQAVETAQAQNGVAMSSLSQSLGSS 299
QY 297 GSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERPGQMIG 356
DB 300 GSSGLGAGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERPGHMIG 359
QY 357 GLPVGQMGARAG--GGTSGVLRVPRPVVPHSPAG 391
DB 360 GLPLGQLTSGGGGVSNALRMPPRVAVPRVPAAG 396

RESULT 2
YS92 MYCTU
ID _YS92 MYCTU STANDARD; PRT; 408 AA.
AC Q10813;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV2892C/MT2959/Mb2916c.
GN RV2892C OR MT2959 OR MFCY274.23C OR MB2916C.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RN SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence."
RL Nature 393:537-544 (1998).
[2]
RN SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=2226494; PubMed=12218036;
RA Fleischmann R.D., Alland A., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,

```

```

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains."
RL J. Bacteriol. 184:5479-5490 (2002).
RN [3]
RN SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Dutjoi S., Grondin S., Lacroix C., Konsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; Z74024; CAA98377.1; -
CC EMBL; AE007119; AAK47285.1; -
CC EMBL; EX248344; CAD96603.1; -
CC FIC; G70925; G70925.
CC TIGR; MT2959; -
CC TubercuList; RV2892C; -
CC InterPro; IPR00030; Microbac_PPE.
CC Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 56 76 POTENTIAL.
SQ SEQUENCE 408 AA; 41469 MW; 3E3D1F20D7827199 CRC64;

Query Match      36.2%; Score 705; DB 1; Length 408;
Best Local Similarity 41.8%; Pred. No. 2.7e-34;
Matches 182; Conservative 47; Mismatches 134; Indels 72; Gaps 14;

QY 2 VDFGALPPEINSMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIGS 61
DB 1 MDFGVLPEINSGMYAGPGSGPMMAAAAWDSLAELGLAAGGYRLAISLGTGAYWAGP 60
QY 62 SAGLMVAASPYVAMSVTAQAGELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 121
DB 61 AAASMVAAVTPYVAVLSATAGQAQAGQAARAAAAYELAFAMTVPPVIVANRALLVAL 120
QY 122 IATNLGQNTPAIAVNEAEYGEWMAQDAAMFCAATATATATLTPFEAPEMTSAGGL 181
DB 121 VATNPFQNTPAIAATEAQVAEMWQAQDAAMAYAGSAATAT-ELTPTTAAPVTTSPAAL 179
QY 182 LEQAAA-----VEEASDTAAANQLMN-----VP--QAQQ--LAQQTGTTTSSKL--- 224
DB 180 AQQAATVSSVTPPLATTAAPVQLQLQSLTSLIPWYSAQQQWLAENLLGTPPNRMTIV 239
QY 225 -----GGLWKTVPVHRSPISNMVSMNNHSMNTNSGVSMTNTLSMLKGFAPAAAA 275
DB 240 RLIGISYFDEGL-----LQFEASLAQAQAPGTPGGAG--DSGSSVLDMSGPTIFA 287
QY 276 QAVTAQNGVAMSSL--GSSLSG-----SCLGGVAAANLGRAASVGSLS 319
DB 288 -----GPRSPSVAGGAVGVQTPQYVWALDRESIGGSVAAALKGSSAGSLS 338
QY 320 VPQAWAANQAVTPAARALP---LTSITSAERPGQMIGLPGVQMGARAGGSLGVLR 376
DB 339 VPDWAARARWANPAANWELPGDDVTALRGTAENA--LLRGFFMASAGQSTGGGF--VHK 393
QY 377 VPPRPYVPHSPAG 391
DB 394 YGFLAVNQRPFPAG 408

```

```
RESULT 3
YI02 MYCTU STANDARD; PRT; 463 AA.
AC O53951;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV1802/MT1851/MD1830.
GN RV1802 OR MT1851 OR MTV049.24 OR MB1830.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
SEQUENCE FROM N.A.
RP SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Horsby T., Jagels K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
SEQUENCE FROM N.A.
RP SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uitterlinden T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490 (2002).
RN [3]
SEQUENCE FROM N.A.
RP SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duchoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -I- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; AL022021; CAAL1723.1; -
CC ENBL; AE007044; AAK46123.1; -
CC ENBL; BX248340; CAD94533.1; -
CC PIR; C70931; C70931.
CC TIGR; MT1851; -
CC TubercuList; RV1802; -
CC InterPro; IPR000030; Microbac_PPE.
CC Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 401 401 S -> I (IN REF. 2).
```

```
SQ SEQUENCE 463 AA; 46021 MW; EE64828BF09FA551 CRC64;
Query Match 35.3%; Score 688.5; DB 1; Length 463;
Best Local Similarity 42.8%; Pred. No. 2.8e-33;
Matches 174; Conservative 48; Mismatches 136; Indels 49; Gaps 11;

QY 2 VDFCALPPEINSARMYAGPGSASLVAAQWDSVADLFSAAAFQSVWGLTVGWSIGS 61
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MDFGVLPPEINSGRMYAGPGSGPMLAAAAAAMDGLATELQSTAAADYGSVISLT- 59
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 62 SAGLMVAAAPYVAVMSVTGAQELTAQVRVAAAAYETAYGLTVPPVIAENPAELMIL 121
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 SSGTMAAAAPYVAVMSGATLAAREAAAAQASAAAAAYEAPATVPPVVAANRAELAVL 119
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 122 IATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPETSAGGL 181
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 120 AATNIFGONTGAIATAAEARYAEWMAQDAAMFGYAGSSVAT-QVTFFAAPPTTNAAGL 178
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 182 LEQAAVEEASDTAAANQLMNVPAQLQOLAQPTQGTTPSKLGLKWTYS--PHRSP1- 238
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 179 ATQGVAVAAQVAGSAGN-ARSLVSEVLEFLA--TAGTNYNKTVASLMNAVTVGVPIASSVY 235
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 239 -----SNMYSMANNHMSNTSGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRA 288
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 236 NSMTGLGFAESKMWLPANDTVISTIFGMVQFQKEFNVTFNPDLPK----- 283
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 289 MSSLGSLSIG-----SSGLGG--GVAANLGRPAASVGLSVDPQAWAANQAVTPAARALPL 340
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 294 -SALGAGLGLRSATSSGLGSTAPASAGASQAGSVGMSVPPSWAAATPAIKTVAAVFSS 342
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 341 TSLTS--AAERGPQGL-----GGLPYGQMGARAGGSLGVLRV 377
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 343 TGLQAVPAAATSEGLSLSQMALASVAGGALGGAARATGGFLGGGRV 389
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 4
Y442 MYCTU STANDARD; PRT; 487 AA.
ID Y442 MYCTU
AC P42611; OS3727;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV0442c/MT0458.
GN RV0442C OR MT0458 OR MTV037.05C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=Erdmann;
RX MEDLINE=87137260; PubMed=3029018;
RA Shinnick T.M.;
RT "The 65-kilodalton antigen of Mycobacterium tuberculosis."
RL J. Bacteriol. 169:1080-1088 (1987).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.B. III, Tekala P.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsby T., Jagels K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [3]
SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / Oshkosh;
```

```

RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.D.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15467; AAA88235.1; ALT INIT.
DR EMBL; AL021932; CAA17399.1; -.
DR EMBL; AE006948; AAK44681.1; -.
DR PIR; C70830; C70830.
DR TIGR; MT0458; -.
DR TubercuList; Rv0442c; -.
DR InterPro; IPR00030; Microbac_PPE.
DR InterPro; IPR002989; Mycobac_Pentapep.
DR Pfam; PF01469; Pentapeptide_2; 5.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 40 40 E -> K (IN REF. 2).
FT CONFLICT 96 96 I -> T (IN REF. 1).
FT CONFLICT 211 211 G -> GNNNIG (IN REF. 1).
FT CONFLICT 487 487 AA; 47247 MW; 97234D5B316C8C7F CRC64;
SQ SEQUENCE 487 AA; 47247 MW; 97234D5B316C8C7F CRC64;

Query Match 22.8%; Score 444; DB 1; Length 487;
Best Local Similarity 32.1%; Pred. No. 4.8e-19; Indels 58; Gaps 12;
Matches 135; Conservative 48; Mismatches 180;

Qy 4 FGALPPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWICSSA 63
Db 6 FAWLPPEINSGALMPAGPGSLIAAATAWGELAEELLASLASLGSVTSLETSGAWLGPSA 65
Qy 64 GLXVAASPVVWMSVTAGQAEHTAAQVRVAAAYETAYGLTVPPPPVIAENRAELMILIA 123
Db 66 AAMVAVATQYLAWLSTAAQAEQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 125
Qy 124 TNLIGQNTPAIVNVEAEYGEWQAQDAAMFGYAAATATATATATLLPPEAPEMTSAG--- 179
Db 126 TNWFGQNALPMQVDEAAEQWALDVAAVAGYHFDASAAVAQLAPHQOV--LRNLGIDIG 183
Qy 180 --GLLEQAAVAVERASDTAAANQLMNNVPQALQLOAQPTQCTTSSKLG-----GLW 228
Db 184 KNGQINLGFNGTSGNIGNNNIGNNIGNTGTGNTGNTGNTGNTGNTGNTGNTGNTGNTGNTG 243
Qy 229 KTVS-----PHRSPISNMVSMANNHSMNTSGVSMNTLSSMLKGFAFAAAQAQVQ 279
Db 244 NTSGNIGFQIGDGHQMGFGFGNSGSGN-IGFNGSGTGNVGLFNS----- 287
Qy 280 TAAQNGVWAMSLGSSGLGGVAAANLGRVAASVGLSPVQAAWAAANQAVTPAARALP 339
Db 288 GSGNIGTSGNSLNGTSGT---INAGLG---SAGSLNT-SFWNAGNQNAALGSAAGS 340
Qy 340 LTSITSAERPGQM-----LGLPLVQMGARAG--GGLSGVLR--VPRPVPYMPHSPAA 390
Db 341 EALVSSAGVATGGMSTAAALSSGILASALGSTGGLQHLANVLNLSGLTNTVPAAPASAV 400
Qy 391 G 391
Db 401 G 401

```

```

Y878 MYCTU STANDARD; PRT; 443 AA.
ID Y878 MYCTU
AC Q10540;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rv0878c/MT0901.
DE Rv0878c OR MT0901 OR MTCY31.06C.
GN Mycobacterium tuberculosis.
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1773;
RN [1] RN
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98293987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2] RN
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol 184:5479-5490(2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z73101; CAA97385.1; -.
DR EMBL; AE006977; AAK45143.1; ALT_INIT.
DR PIR; C70780; C70780.
DR TIGR; MT0901; -.
DR TubercuList; Rv0878c; -.
DR InterPro; IPR00030; Microbac_PPE.
DR InterPro; IPR002989; Mycobac_Pentapep.
DR Pfam; PF01469; Pentapeptide_2; 4.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Transmembrane; Repeat; Complete proteome.
FT TRANSMEM 15 35 POTENTIAL.
FT TRANSMEM 38 58 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 181 201 POTENTIAL.
FT DOMAIN 64 73 POLY-ALA.
FT DOMAIN 81 115 ALA-RICH.
FT DOMAIN 231 270 4 X 10 AA APPROXIMATE REPEATS.
SQ SEQUENCE 443 AA; 43592 MW; C58BEC607F0675E2 CRC64;

Query Match 21.9%; Score 426.5; DB 1; Length 443;
Best Local Similarity 31.2%; Pred. No. 4.5e-18;
Matches 125; Conservative 56; Mismatches 132; Indels 87; Gaps 14;

```


QY	2	VDFGALPPEINSA	RYAGFGSASLVAAQAQMWDSVASOLFSAASAPOSVWGLTVG	---	SW	58
		:::	:::	:::	:::	:::
Db	1	MNFVLPPEVNSA	RYAGAGPAPMLAAAVAWDGLAABLGMWAAAFSLILSGLTAGFGSAW	60		
		:::	:::	:::	:::	:::
QY	59	IGSSAGLVAAAS	PVVAWMSVTAGQELTAAQVNRVAAAAYETAYGLTVPPVIAENRAEL	118		
		:::	:::	:::	:::	:::
Db	61	QCPAAAAA	AAAAAPVLSWLNATARAECAGAKAAAAVTEARAATAHPALVAANNQOL	120		
		:::	:::	:::	:::	:::
QY	119	MILIATNL	LGQNTPAIAVNEABYGEWMAQDAAMFGYAAATATATATLIPFEEAPEMSTA	178		
		:::	:::	:::	:::	:::
Db	121	LSLVLSNLF	QGNQLPALAATAEASVEQLWAQDVAAMVGVHGGASTVASQLTPWQO	---	173	
		:::	:::	:::	:::	:::
QY	179	GGLLQCAAA	VEEASDTAAANQMNVPQALQQLAOPQTGGTTPSSKGLGLWKTVPSPHRPI	238		
		:::	:::	:::	:::	:::
Db	174	---	LLSVLPVTTAA	PAGAV-----GVPA-----LAIPALGV-----ENIG-----	---	208
		:::	:::	:::	:::	:::
QY	239	SNMV---	SNVANNHMTNSG-----VSMTN-----TLSSMLKGFAPAAAAQAV	278		
		:::	:::	:::	:::	:::
Db	209	GNFLGIGN	IGNNVVSGNTGDNFGEIGNIGNANLGNIGNANLGSNAGFFNGNGNDG	268		
		:::	:::	:::	:::	:::
QY	279	QTAQNGV	RANSSLOS--SLGSSGLGGVAA--NLGRAASVGSLSVCAWAAANQVTPAA	335		
		:::	:::	:::	:::	:::
Db	269	NTNFGSG	NAGFLINISGNEGSGNLGFGNAGDDNTG-----WNSGD-----	---	309	
		:::	:::	:::	:::	:::
QY	336	RALPLT	LSITSAAE	RPQGMQLGGL--PVQMGARAGGGILSG	373	
		:::	:::	:::	:::	:::
Db	310	-----	TNTGGFN	SGDLNTIGISPTQGVANSFGCNTG	341	
		:::	:::	:::	:::	:::

RESULT 6

YF48 MYCTU

STANDARD;

PRT;

678 AA.

ID

YF48 MYCTU

AC

Q10778;

DT

01-OCT-1996

(Rel. 34, Created)

DT

15-JUL-1999

(Rel. 38, Last sequence update)

DT

10-OCT-2003

(Rel. 42, Last annotation update)

DE

Hypotheetical PPE-family protein RV1548C/MTI599.

GN

RV1548C OR MTI599 OR MTCY48.17.

OS

Mycobacterium tuberculosis.

OC

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC

Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX

NCBI_taxid=1773;

RN

[1]

RP

SEQUENCE FROM N.A.

RC

STRAIN=H37Rv;

RC

MEDLINE=98295987; PubMed=96314230;

RA

Cole S.T., Bresch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA

Gordon S.V., Eiglmeier K., Gas S., Barry C.E.III, Tekala F.,

RA

Baddock K., Basham D., Brown D., Chillingworth I., Connor R.,

RA

Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA

Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA

Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA

Rutter S., Seeger K., Skelton S., Squares S., Squares R.,

RA

Sulston J.E., Taylor K., Whitehead S., Barrell B.G.

RT

"Deciphering the biology of Mycobacterium tuberculosis from the

RT

complete genome sequence.";

RL

Nature 393:537-544(1998)."

RN

[2]

RP

SEQUENCE FROM N.A.

RC

STRAIN=CDC 1551 / Oshkosh;

RC

MEDLINE=22206494; PubMed=12218036;

RA

Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

RA

Peterson J.J., DeBoy R., Dodson R., Gwinn M., Hart D., Hickey E.,

RA

Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,

RA

Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,

RA

Bisai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.

RT

"Whole-genome comparison of Mycobacterium tuberculosis clinical and

RT

laboratory strains".

RL

J. Bacteriol. 184:5479-5490(2002).

CC

-!- SIMILARITY: Belongs to the mycobacterial pPE family.

CC

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC

between the Swiss Institute of Bioinformatics and the EMBL outstation

```

RX MEDLINE=95020554; PubMed=7934845;
RA Rinke de Wit T.F., Clark-Curtiss J.E., Abebe F., Kolkon A.H.J.,
RA Jonson A.A.M., Thole J.E.R.;
RT "A Mycobacterium leprae-specific gene encoding an immunologically
RT recognized 45 kDa protein."
RL Mol. Microbiol. 10:829-838(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser J., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
RC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U00015; AAC43220.1; -
DR EMBL; X68431; CAA48480.1; -
DR EMBL; Z21952; CAA79950.1; -
DR EMBL; Z97179; CAB09938.1; -
DR EMBL; AL583918; CAC29919.1; -
DR PIR; C86960; C86960.
DR PIR; S33522; S33522.
DR PIR; S39872; S39872.
DR Leproma; ML0411; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Antigen; Repeat; Complete proteome.
FT DOMAIN 192 196 POLY-SER.
FT DOMAIN 209 235 2 X 6 AA REPEATS OF S-V-A-Q-S-E.
FT REPEAT 209 214 1.
FT REPEAT 230 235 2.
FT CONFLICT 132 132 T -> S (IN REF. 2).
FT CONFLICT 189 189 S -> L (IN REF. 2).
FT CONFLICT 191 191 H -> D (IN REF. 2).
FT CONFLICT 292 292 P -> L (IN REF. 2).
SQ SEQUENCE 408 AA; 42466 MW; 5C0C2BE0D6BEA9D8 CRC64;

Query Match
Best Local Similarity 26.3%; Pred. No. 1.7e-14;
Matches 115; Conservative 72; Mismatches 163; Indels 87; Gaps 11;

Qy 1 MVDGALPPEINSARMYAGGCSASLVAAQWDSVASDLFSASAFQSVWGLTVGSWIG 60
Db 1 MFDWYSPVFNAPFLMSRGPGSTPLWGAENAWISLAQLMENAQEVSDTIWVAVPASFAG 60
Qy 61 SSAGLMAAASPYVAMSVTAGQALTAQVRAAAAYETAYGLTVPVPIANRAELMI 120
Db 61 ETSMDLASRSTVFVWLDGNAENAGLTARVHLHAVAYAFEERAGVPLLTVLGNIIHTWA 120
Qy 121 LIATNLGQNTPAJAVNEAYGENWAGQAAAPGYYAAATATATATLTPPEAPEMTSAGG 180
Db 121 LKAINWFGQVSTTVAALEADYDLMMVQNSTANTTYRDLVLRGTGKMFEPAPQLVSR-Y 179
Qy 181 LLEQAAAEASDPTAAANQLMNVPQALQLOAQ----- 213
Db 180 CMERDVSNSFHSSSSSDSLYESIDNLYDSVAQSEERHSGDSMSQSYNTCGSVAQSELCD 239

```

```

QY 214 ----PTQ-----GTTPTSSKLGGLWKIVSPHRSPISNMYSMANNHMTNMGVSMINT 261
DB 240 PGTFSQSSQNDLSATSLTQQLGGL-----DSIISASASALLTINS--ISSST 286
QY 262 LSSMLKGFAPAAAAQAVTAQAQNGVYRAMSSSLGSSGLGGGVAAN-----LGRASAVG 316
DB 287 ASSIM-----PIVASQVTTGLRSQV-AVEKMTQSTSTAVSVDVAASKVAVGQAVSVG 341
QY 317 SLSVPCAAWAAQAQVTPAARALP--LTSILTSAAERGPQGLGLPVGQMGARAGGGLSV 374
DB 342 ALRVENWATASQPVNATAHSPAGCSAITTA-----VSGPLEGV 381
QY 375 LRVPPRPVYVMPHSPAAG 391
DB 382 TQ--PAEEVLITASVAGG 396
RESULT 8
Y096_MYCTU
ID Y096_MYCTU STANDARD; PRT; 463 AA.
AC Q10892;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE Hypothetical PPE-family protein Rv0096/MT0105.
GN Rv0096 OR MT0105 OR MTCY251.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OC NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z74410; CAA98932.1; -
DR EMBL; AE006922; AAK44327.1; -
DR PIR; H70750; H70750.
DR TIGR; MT0105; -
DR TubercuList; Rv0096; -

```

```
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 216 236 POTENTIAL.
FT TRANSMEM 245 265 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 419 439 POTENTIAL.
SQ SEQUENCE 463 AA; 429D9D66A033D0D8 CRC64;

Query Match
Best Local Similarity 16.6%; Score 324.5; DB 1; Length 463;
Matches 112; Conservative 54; Mismatches 144; Indels 93; Gaps 12;

QY 6 ALPPEINARMYAGPGSGASLVAAQWDSVASDLFSAASAFQSVWGLTVGSGWIGSAGL 65
DB 2 AIPEVHSGLLSAGCGPGSLVAAQWQELSDQYALACAEGLGQLLGEVQASSWQGTAA 61
QY 66 MYAASPYVAMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIATN 125
DB 62 YVAAHGPEYLAWELETAINSVTAQHVAAAAYCSALAAFTPAELAAAHAIHGVLIATN 121
QY 126 LLGONTPTAIVNEAEYGEWMAQDAAMFGVAAATATATATALLPPEAPEMTSAGLL 185
DB 122 FFGINTVPIALNEADYVRMWLQADTNAALQVADAATVAVPTQAPPIRAPGG----- 176
QY 186 AAVEEASDT-----AAANQLNMVPAQLOOLAQPTQ-----GTPSSKL----- 224
DB 177 ----DAADTRLVLSIGLIRDI---LDFIANPYKVFLEFFQFGFSPAVTVVVALVAL 229
QY 225 ---GGLW-----KTVSHRSPISNMVSMANNHSMVNTNLSMLK-----GF 269
DB 230 QLVDFLWYPYVYAGVGLLLPFFTP-----TLSATLALHLLNLPPAGL 275
QY 270 APAQAQAVQTAQNGVRAMSSLCSSGLGGVAAANLGRAASVGSLSVPOAWAAANQ 329
DB 276 LPIAAA-----LPGDQWGANLAVATATAVP-----GSGP 308
QY 330 AVTPAPRALPLTSLTSAERGGP---QMLGGLPVG-OMGARAG 368
DB 309 PTPNPAPAPSPSSVGSASAPGISYAVPLAPPVSGSPKAG 351

RESULT 9
YU18_MYCTU STANDARD; PRT; 434 AA.
AC P31500; O53265;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV3018C/MT3098/MT3101.
GN RV3018C OR MT3098/MT3101 OR MTV012.32C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cele S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
```

298 GGPLLGAALAAVPGVAGLAGVLAAL-PAVGAA--AGAPAAALVGSVPVSGVSPQA 354
 324 WAAANQAVTPAARALPILTSLSAERPGQMLGGLPVGQWAGAGGSLGV 374
 355 RLVS--AVEPAPASTSVSL--ASDRGAGAL--GF-VGTAGKESVGGPAGL 398

RESULT 10
 YU21 MYCTU
 ID YU21 MYCTU STANDARD; PRT; 435 AA.
 AC 053268; OS3269;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein RV3021C/RV3022C/MT3106.
 GN RV3021C/RV3022C OR MT3106 OR MTV012.35C/MTV012.36C.
 OS Mycobacterium tuberculosis
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Teklaia F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin S., Holroyd S.,
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains";
 RL J. Bacteriol. 184:5479-5490(2002).
 CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
 frameshift in position 82.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AL021287; CAAL16106.1; ALT_FRAME.
 DR EMBL; AL021287; CAAL16107.1; ALT_FRAME.
 DR EMBL; AE007129; AAK47435.1; -.
 DR TIGR; MT3106; -.
 DR TubercuList; RV3021C; -.
 DR TubercuList; RV3022C; -.
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 DR Hypothetical protein; Complete proteome.
 KW CONFLICT 299 G -> A (IN REF. 2).
 FT CONFLICT 317 LAGV -> VTGL (IN REF. 2).
 FT CONFLICT 326 L -> V (IN REF. 2).
 SQ SEQUENCE 435 AA; 42876 MW; 38157643EAA8484A CRC64;

Query Match 16.5%; Score 321.5; DB 1; Length 435;
 Best Local Similarity 26.5%; Pred. No. 5.6e-12;
 Matches 115; Conservative 53; Mismatches 15; Indels 107; Gaps 12;

QY 6 ALPPEINARMVAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSGWIGSSAGL 65
 DB 8 ASPPEVHSALLSAGPGPSLQAAAGWSALSAAEVAQAQELSVAQVAVAGAGVWQGPSAEL 67
 QY 66 MVAAAAPYVAMNSVTAGQAEELTAQVRVAAAYETAYGLTVPPPPVIAENRAELMILITATN 125
 DB 68 FVAAYVPVAVMLVQASADSAASAAAGEHEAAAGYVCAALAEEMPTLPELAANHLTHAVLVATN 127
 QY 126 LLGQNTPAIVNEAYGEMWQAQDAAMAFGYAA----- 157
 DB 128 FFGINTIFIALNEADYVRWVQAATVMSAYEAVVGAALVATPHTGPPAVIVKPGANEASN 187
 QY 158 AVATATATALLPPEE-----APEMTSAGGLLEQAAAEEASDTAAANQIMNNVPOALQ 209
 DB 188 AVAAATITPPFPFGLAKLEMAAQAFTEVGEILMKSAEAWAVGFVELITGLVNFEP--- 243
 QY 210 QLAQPTQTPTSSKLGGLWKVSPHRSPISNVSM-----ANNH 248
 DB 244 -----WLW-----LTGMIDMFFATVGFALGVFLVPLLEFAVVLE 278
 QY 249 MSMTNSGVSMNTLSS-----MLKGFAPAAAQAQVOTAAQNGVYRAMSSLSLSSGSL 301
 DB 279 LAILIGHIISNFGAIPVLGGPLLGAALAAVPGVAGLAGVLAALPAVGAAGAP-- 336
 QY 302 GCGVAANTGRASV--GSLSVPOQMAAANQAVTPAARALPLTSLTSAERPGQMLGGLPV 360
 DB 337 -----AALVGSVAPVSGGVVSPQARLVS--AVEPAPASTSVSL--ASDRGAGAL--GF-V 385
 QY 361 GQMGARAGGSLGV 374
 DB 386 GTAGKESVGGPAGL 399

RESULT 11
 YU29 MYCTU
 ID YU29 MYCTU STANDARD; PRT; 178 AA.
 AC 006246;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein RV3429/MT3533.
 GN RV3429 OR MT3533 OR MTCV77.01.
 OS Mycobacterium tuberculosis
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Teklaia F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin S., Holroyd S.,
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains";
 RL J. Bacteriol. 184:5479-5490(2002).
 CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
 frameshift in position 82.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AL021287; CAAL16106.1; ALT_FRAME.
 DR EMBL; AL021287; CAAL16107.1; ALT_FRAME.
 DR EMBL; AE007129; AAK47435.1; -.
 DR TIGR; MT3106; -.
 DR TubercuList; RV3021C; -.
 DR TubercuList; RV3022C; -.
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 DR Hypothetical protein; Complete proteome.
 KW CONFLICT 299 G -> A (IN REF. 2).
 FT CONFLICT 317 LAGV -> VTGL (IN REF. 2).
 FT CONFLICT 326 L -> V (IN REF. 2).
 SQ SEQUENCE 435 AA; 42876 MW; 38157643EAA8484A CRC64;

```

CC -I- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; Z77165; CAB01031.1; -.
CC DR PIR; F70738; F70738.
CC DR TubercuList; Rv3425; -.
CC DR InterPro; IPR000030; Microbac_PPE.
CC DR Fram; PF00823; PPE; 1.
CC KW Hypothetical protein; Complete proteome.
CC SQ SEQUENCE 176 AA; 19855 MW; B8CF2B9463B87B0 CRC64;
CC -----
Query Match 11.2%; Score 217.5; DB 1; Length 176;
Best Local Similarity 33.5%; Pred. No. 2.3e-06;
Matches 58; Conservative 31; Mismatches 83; Indels 1; Gaps 1;
CC -----
QY 7 LPPEINARWYAGFGASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGSAGLM 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 IPAEYISNIIEGFGADSLFFASGQLRELAYSVETTAESLDELDLDELDELDELDEL 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 VAASPVPVAMSVYTAQAEITAAQVRVAAAYETAYGLTVPPVIAENRAELMILITATNL 126
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 ADAVERYLQWLSKHSQLKHAANVINGLANAYNDRKVVPPPEILANREERRLIASNV 123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 127 LGQNTPAIAVNEAEYGEWMAQDAAMVFGAAATATATATATALLPPEEAPMTSAG 179
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 AGVNTTPAIDLDAQYQYQYARNVAVNNVYVSVTRSAISLDPWRREPPQIYRGG 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 13
ELS_MOUSE STANDARD; PPT; 860 AA.
AC P54320;
RC 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Elastin precursor (tropoelastin).
ES ELN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Lung;
RX MEDLINE=95130069; PubMed=7829060;
RA Wyder K.S., Sechler J.L., Boyd C.D., Passmore H.C.;
RT "Use of an intron polymorphism to localize the tropoelastin gene to
RT mouse chromosome 5 in a region of linkage conservation with human
RT chromosome 7.";
RL Genomics 23:125-131(1994).
CC -I- FUNCTION: Major structural protein of tissues such as aorta and
CC nuclear ligament, which must expand rapidly and recover completely.
CC -I- SUBUNIT: The polymeric elastin chains are cross-linked together
CC into an extensible 3D network.
CC -I- SUBCELLULAR LOCATION: Extracellular matrix of elastic fibers.
CC -I- PM: The crosslinks are made of deaminated Lys.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; U08210; AAA80155.1; -.

```

```

DR PIR; A55721; EAMS.
DR MGD; MGI:95317; Eln.
DR InterPro; IPR003979; tropoelastin.
DR PRINTS; PRO1500; TROPOLASTIN.
KW Structural protein; Repeat; Signal; Connective tissue.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 860 ELASTIN.
FT DISULFID 850 855 BY SIMILARITY.
SQ SEQUENCE 860 AA; 71955 MW; 0C0B55AAE1EDD7F1 CRC64;

Query Match 7.8%; Score 153.5; DB 1; Length 860;
Best Local Similarity 24.4%; Pred. No. 0.068;
Matches 115; Conservative 34; Mismatches 192; Indels 131; Gaps 17;

QY 5 GALT-----PEINSMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSW-- 58
DQ 402 GGIPGVGGPGIGGPGIYGGPGCAVSPAAAATAAATAA--YCARGGVGIPYGVGAGGFFG 459
QY 59 --IGSSAGLVAAASPVVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPVIAENRA 116
DQ 460 YGVGAGAGL--GGASPA--AAAATAAAYG-----AGGAG 491
QY 117 ELMILIAITNLGQNTPAIAVNEAYGEMWAQDAAMFGEYAAATATATATLTP-FEBAPEM 175
DQ 492 ALGLGVPGAGVPGALPGAVPAGVPGAGTFAAAAAATAAATAAAGAGLPGVGGVPGG 551
QY 176 TSAGGL-----LEQAAVEEASDFTAANQ----- 199
DQ 552 VGVGGIPGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGGVGG 611
QY 200 LMNVPPQALQLAQTOGT-----TPSSKLGGLWKTVPSPHRSISINNVSMANN 247
DQ 612 LGAGVPGFAGAGVPGFAGAGVPGFAGAGVPGFAGAGVPGFAGAGVPGFAGAGVPGFAG 671
QY 248 HSMSTN-----SGVSMNTNLSMLKGFAPAAAQAAQVTAQNGVRA 288
DQ 672 YGAGAGLGGPGGLGGPGGLGGPGGLGGPGGLGGPGGLGGPGGLGGPGGLGGPGGLGG 731
QY 289 MSSLGS--SLGSSGLG--GVVAANLGRASVGSLSVPQAAANQAVTPAARALPLTSLTSA 346
DQ 732 AGGLGAGLGGAGLGGAGLGGAGLGGAGLGGAGLGGAGLGGAGLGGAGLGGAGLGGAGL 784
QY 347 AERGP-----GOMLGG--LPVQOMGARAGGLS-----GVLRVPPRPY 382
DQ 785 AKYGAAGLGGVILGARPPFGGVAARPGFGLSPYVGGGAGLGGVGGKPKPKPY 836

RESULT 14
YI26 MYCTU STANDARD; PRT; 232 AA.
AC Q50702;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PPE-family protein RV3426.
GN RV3426 OR MTCY78.03C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98255987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagals K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL; Z77165; CAB01030.1; -.
DR PIR; G70738; G70738.
DR Tuberculist; RV3426; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 232 AA; 25872 MW; D76512D49EB272C6 CRC64;

Query Match 7.8%; Score 151.5; DB 1; Length 232;
Best Local Similarity 35.2%; Pred. No. 0.021;
Matches 43; Conservative 18; Mismatches 60; Indels 1; Gaps 1;

QY 7 LPPEINSMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSWIGSSAGLM 66
DQ 5 IPAEVISNVIVGPRADSLYAADQRLQADSVRTTAESLNTLDELH-ENWKGSSSEWM 63
QY 67 VAAASPVVAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPVIAENRAELMILIAITNL 126
DQ 64 ADAALRYLDLMSKHSRQLRTARVIESLVMAEETLLRVVPPPATIANRREVRLLIASNV 123
QY 127 LG 128
DQ 124 AG 125

RESULT 15
EIS RAT
ID ELS RAT STANDARD; PRT; 864 AA.
AC Q99372;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Elastin precursor (Tropoelastin) (Fragment).
GN ELN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=91104868; PubMed=1702999;
RA Pierce R.A., Deak S.B., Stolle C.A., Boyd C.D.;
RT "Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.";
RL Biochemistry 29:9677-9683(1990).
RN [2]
RP SEQUENCE OF 781-864 FROM N.A.
RX MEDLINE=86330868; PubMed=2971041;
RA Deak S.B., Pierce R.A., Belsky S.A., Riley D.J., Boyd C.D.;
RT "Rat tropoelastin is synthesized from a 3.5-kilobase mRNA.";
RL J. Biol. Chem. 263:13504-13507(1988).
RN [3]
RP SEQUENCE OF 264-533 AND 558-864 FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=92241859; PubMed=1572837;
RA Pierce R.A., Alatawi A., Deak S.B., Boyd C.D.;
RT "Elements of the rat tropoelastin gene associated with alternative
RT splicing.";
RL Genomics 12:651-658(1992).
CC -!- FUNCTION: Major structural protein of tissues such as aorta and
CC nuchal ligament, which must expand rapidly and recover completely.
CC -!- SUBUNIT: The polymeric elastin chains are cross-linked together
CC into an extensible 3D network.

```

Search completed: June 22, 2004, 17:16:54
Job time : 7.59814 secs

```
CC CC -!- SUBCELLULAR LOCATION: Extracellular matrix of elastic fibers.
CC CC -!- ALTERNATIVE PRODUCTS:
CC CC Event=Alternative splicing; Named isoforms=8;
CC CC Comment=Experimental confirmation may be lacking for some
CC CC isoforms;
CC CC Name=1;
CC CC IsoId-Q99372-1; Sequence=Displayed;
CC CC Name=2;
CC CC IsoId-Q99372-2; Sequence=VSP_004244;
CC CC Name=3;
CC CC IsoId-Q99372-3; Sequence=VSP_004245;
CC CC Name=4;
CC CC IsoId-Q99372-4; Sequence=VSP_004246;
CC CC Name=5;
CC CC IsoId-Q99372-5; Sequence=VSP_004244, VSP_004245;
CC CC Name=6;
CC CC IsoId-Q99372-6; Sequence=VSP_004245, VSP_004246;
CC CC Name=7;
CC CC IsoId-Q99372-7; Sequence=VSP_004244, VSP_004246;
CC CC Name=8;
CC CC IsoId-Q99372-8; Sequence=VSP_004244, VSP_004245, VSP_004246;
CC CC -!- PTM: The crosslinks are made of deaminated lys.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC at the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to licenses@isb-sib.ch).
-----
CC CC EMBL; M60647; AAA42269.1; --
CC CC EMBL; J04035; AAA42268.1; --
CC CC EMBL; M86372; AAA42271.1; --
CC CC EMBL; M86355; AAA42271.1; JOINED.
CC CC EMBL; M86363; AAA42271.1; JOINED.
CC CC EMBL; M86364; AAA42271.1; JOINED.
CC CC EMBL; M86366; AAA42271.1; JOINED.
CC CC EMBL; M86371; AAA42271.1; JOINED.
CC CC EMBL; M86376; AAA42272.1; --
CC CC EMBL; M86373; AAA42272.1; JOINED.
CC CC EMBL; M86375; AAA42272.1; JOINED.
CC CC PIR; A36106; EART.
CC CC InterPro; IPR003979; tropoelastin.
CC CC PRINTS; PR01500; TROPOELASTIN.
CC CC Structural protein; Connective tissue; Repeat; Signal;
CC CC Alternative splicing.
CC CC FT NON_TER 1
CC CC FT SIGNAL <1 21 BY SIMILARITY.
CC CC FT CHAIN 22 864 ELASTIN.
CC CC FT DISULFID 854 859 BY SIMILARITY.
CC CC FT VARSPPLIC 263 307 Missing (in isoform 2, isoform 5, isoform
CC CC FT 7 and isoform 8). /FTid=VSP_004244.
CC CC FT VARSPPLIC 308 308 Missing (in isoform 3, isoform 5, isoform
CC CC FT 6 and isoform 8). /FTid=VSP_004245.
CC CC FT VARSPPLIC 809 823 Missing (in isoform 4, isoform 6, isoform
CC CC FT 7 and isoform 8). /FTid=VSP_004246.
CC CC SEQUENCE 864 AA; 72786 MW; 456894BE09E79D4 CRC64;
-----
Query Match 7.7%; Score 150; DB 1; Length 864;
Best Local Similarity 25.9%; Pred.No. 0.11;
Matches 102; Conservative 31; Mismatches 153; Indels 108; Gaps 18;
-----
QY 5 GALP----PEINSARMYAGPGSLVAQAQMWDVSADLSFAASAFOSVVWGLTVGSW-- 58
Db 399 GGIPGVGGPIGGPIGVGGPGAVSPAATAAKAAKAAK--YGARGGVGIPYGVGAGGPPG 456
QY 59 --IGSSAGLMVAASPYPVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIENRA 116
Db 457 YGVGAGAGL-----GGASAQAAAA---AAAKAAKYAG----- 486
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:55:54 ; Search time 31.9703 Seconds

(without alignments)
3858.816 Million cell updates/sec

Title: US-09-886-349A-14

Perfect score: 1949

Sequence: 1 MVDFGALPEINSARMYAGP.....SGVLVPPRPYVWPHSPAAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25: *

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacterioplasm:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1949	100.0	391	16	005298
2	1920.5	98.5	390	16	Q7U0E9
3	1779.5	91.3	396	16	Q7U071
4	1648	84.6	393	16	Q8VIZ3
5	1583	81.2	393	16	Q06341
6	1583	81.2	393	16	Q7TW55
7	779.5	40.0	393	16	Q7TZJ3
8	779.5	40.0	411	16	Q53939
9	754	38.7	410	2	Q990T1
10	751	38.5	409	16	Q53957
11	749	38.4	409	16	Q7TZH7
12	740.5	38.0	399	16	Q7TZH8
13	737.5	37.8	403	16	Q53956
14	736	37.8	423	16	Q53950
15	736	37.8	423	16	Q7TZI4
16	731	37.5	421	16	Q9Z5K0

17	702	36.0	391	16	P96362
18	702	36.0	391	16	Q7U0T5
19	702	36.0	413	16	Q06386
20	683.5	35.1	694	16	Q8VUW0
21	682.5	35.0	468	16	O53958
22	675	34.6	380	16	P95190
23	675	34.6	380	16	Q7TX66
24	670.5	34.4	385	16	Q7TZ87
25	669.5	34.4	394	16	Q7TXX5
26	668.5	34.3	462	16	Q33310
27	667.5	34.2	385	16	Q33204
28	666.5	34.2	385	16	Q8VJZ0
29	641	32.9	364	16	Q7TZJ5
30	637	32.7	350	16	Q7TZJ2
31	635	32.6	363	16	O53940
32	633.5	32.5	365	16	O86373
33	629.5	32.3	405	16	Q8VJW5
34	626.5	32.1	381	16	Q7TX67
35	618.5	31.7	397	2	Q9AGF0
36	615.5	31.6	382	16	Q7TXX3
37	615.5	31.6	402	16	Q33312
38	610	31.3	443	16	O8VKL9
39	604	31.0	443	16	Q7U242
40	603	30.9	423	16	Q7U114
41	603	30.9	426	16	O05907
42	597.5	30.7	391	16	O05798
43	594.5	30.5	391	16	Q7TX76
44	589	30.2	406	16	P71869
45	589	30.2	406	16	Q7TW99

ALIGNMENTS

RESULT 1

005298 PRELIMINARY; PRT; 391 AA.
 ID C05298
 AC C05298;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Hypothetical protein (PPE family protein).
 GN RV1196 OR MTC1364.08 OR MT1234.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;
 RX MEDLINE=98295387; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Rutter S., Seeger K., Skelton S., Squares R.,
 Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

P96362	mycobacteri
Q7U0T5	mycobacteri
Q06386	mycobacteri
Q8VJW0	mycobacteri
O53958	mycobacteri
P95190	mycobacteri
Q7TX66	mycobacteri
Q7TZ87	mycobacteri
Q7TXX5	mycobacteri
Q33310	mycobacteri
Q33204	mycobacteri
Q8VJZ0	mycobacteri
Q7TZJ5	mycobacteri
Q7TZJ2	mycobacteri
O53940	mycobacteri
O86373	mycobacteri
Q8VJW5	mycobacteri
Q7TX67	mycobacteri
Q9AGF0	mycobacteri
Q7XX3	mycobacteri
O33312	mycobacteri
O8VKL9	mycobacteri
Q7U242	mycobacteri
Q7U114	mycobacteri
O05907	mycobacteri
O05798	mycobacteri
Q7TX76	mycobacteri
P71869	mycobacteri
Q7TW99	mycobacteri


```

QY 181 LLEQAAVEEASDTAAANQNMNNVPOALQOLAQTGTTGTTSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAVEEASDTAAANQNMNNVPOALQOLAQTGTTGTTSSKLGGLWKTVPSPHRSPI 240
QY 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAAAQVTAQAQNGVRAMSS-----LGSSL 296
DB 241 IVSMLNNHVMTSGVSMNTLSSMLKGFAP-AAAQAVETAQAQNGVQAMSSLSGSLGSSL 299
QY 297 GSSGLGGVAAANLGRAASVGSLSVPPQAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 356
DB 300 GSSGLGGVAAANLGRAASVGSLSVPPQAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 359
QY 357 GLPVGQMGARAG--GGLSGVLVPPRPYPVPHSPAAG 391
DB 360 GLPLGQLTNSGCGFGVSNALRMPPRAYVMPVPAAG 396

RESULT 4
ID Q8V123 PRELIMINARY; PRT; 393 AA.
AC Q8V123;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE PPE family protein.
GN MT3582.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bhalai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE007161; AAK47941.1; -
DR TIGR; MT3582; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR PROSITE; PS00501; SPASE I.1; 1.
SQ SEQUENCE 393 AA; 39658 MW; 96F0B67798855511 CRC64;

Query Match 84.6%; Score 1648; DB 16; Length 393;
Best Local Similarity 84.8%; Pred. No. 2e-85;
Matches 335; Conservative 20; Mismatches 34; Indels 6; Gaps 3;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTGVSGWG 60
DB 1 MVDGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTGVSGWG 60
QY 61 SSAGLMVAASAPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
DB 61 SSAGLMVAASAPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
QY 121 LIATNLGQNTPAIAVNEAYEGEMWAQDAAMFGYAAATATATATLLPPEAEPMTSAGG 180
DB 121 LIATNLGQNTPAIAVNEAYEGEMWAQDAAMFGYAAATATATATLLPPEAEPMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQNMNNVPOALQOLAQTGTTGTTSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAVEEASDTAAANQNMNNVPOALQOLAQTGTTGTTSSKLGGLWKTVPSPHRSPI 240
QY 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAAAQVTAQAQNGVRAMSS-----LGSSL 296
DB 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAAAQVTAQAQNGVRAMSS-----LGSSL 296
QY 297 GSSGLGGVAAANLGRAASVGSLSVPPQAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 356
DB 297 GSSGLGGVAAANLGRAASVGSLSVPPQAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 356

```

```

DB 241 IVSMLNNHVMTSGVSMNTLSSMLKGFAP-AAAQAVETAQAQNGVQAMSSLSGSLGSSL 299
QY 297 GSSGLGGVAAANLGRAASVGSLSVPPQAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 356
DB 300 GSSGLGGVAAANLGRAASVGSLSVPPQAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 359
QY 357 GLPVGQMGARAGGGLSGVLVPPRPYPVPHSPAAG 391
DB 360 GLPLGH-SVNAGSGINNALRVPARAYAIPTPTPAAG 393

RESULT 5
ID O06341 PRELIMINARY; PRT; 393 AA.
AC O06341;
DT 01-JUL-1997 (T-EMBLrel. 04, Created)
DT 01-JUL-1997 (T-EMBLrel. 04, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical protein RV3478.
GN RV3478 OR MTCY13E12.31.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Gas S., Barry C.E. III, Tekai F.,
RA Gordon S.V., Biglmeier K., Brown D., Chillingworth T., Connor R.,
RA Badcock K., Basham D., Brown D., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Davies R., Devlin K., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
DR EMBL; Z95390; CAB08702.1; -
DR PIR; C70568; C70568.
DR Tuberculist; RV3478; -
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 81.2%; Score 1583; DB 16; Length 393;
Best Local Similarity 81.5%; Pred. No. 9.3e-82;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTGVSGWG 60
DB 1 MVDGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTGVSGWG 60
QY 61 SSAGLMVAASAPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
DB 61 SSAGLMVAASAPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
QY 121 LIATNLGQNTPAIAVNEAYEGEMWAQDAAMFGYAAATATATATLLPPEAEPMTSAGG 180
DB 121 LIATNLGQNTPAIAVNEAYEGEMWAQDAAMFGYAAATATATATLLPPEAEPMTSAGG 180
QY 181 LLEQAAVEEASDTAAANQNMNNVPOALQOLAQTGTTGTTSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAVEEASDTAAANQNMNNVPOALQOLAQTGTTGTTSSKLGGLWKTVPSPHRSPI 240
QY 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAAAQVTAQAQNGVRAMSS-----LGSSL 296
DB 241 VSSIANNHSMNTGTCVSMNTLHSLMKGLAP-AAAQAVETAQAQNGVQAMSSLSGSLGSSL 299
QY 297 GSSGLGGVAAANLGRAASVGSLSVPPQAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 356

```

```

Db 300 GSSGLGAGVAANLGRAASVGLSVPPAWAANQAVTPPAARALPLTSLTSAQTAPGHMLG 359
QY 357 GLPVQMGARAGGGLSGVLRVPPRPVYMPHSPAAG 391
Db 360 GLPLGH-SVNAAGSGINNLRVPAAYAIPTPAAG 393

RESULT 6
Q7TWFS PRELIMINARY; PRT; 393 AA.
AC Q7TWFS5
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DE PE family protein.
GN PP260 OR MB3505.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97; PubMed=12788972;
RA Garnier T., Eigleier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248346; CAD95692.1; -.
KW Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 81.2%; Score 1583; DB 16; Length 393;
Best Local Similarity 81.5%; Pred. No. 9.3e-82;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 1 MVDGALPPEINSARMYAGPSASLVAAQWDSVADLFSASAFQSVVWGLTVGSWIG 60
Db 1 MVDGALPPEINSARMYAGPSASLVAAQWDSVADLFSASAFQSVVWGLTVGSWIG 60
QY 61 SSAGLMVAASAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLMVAASAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGQNTPAIAVNEAYEGEMWADAAAFYAAATATATATLPPPEAPEMTSAGG 180
QY 121 LIATNLLGQNTPAIAVNEAYEGEMWADAAAFYAAATATATATLPPPEAPEMTSAGG 180
QY 181 LLEQAAVEEASDIAIDTAANQLMNNVPOALQQLAQPAQWVPSKLGGLWTAVALSPLSN 240
QY 181 LLEQAAVEEASDIAIDTAANQLMNNVPOALQQLAQPAQWVPSKLGGLWTAVALSPLSN 240
QY 241 MVSMANNHMTSGVSMNTLSMLKGFAPAAAAQVQTAAGNGVAMSS----LGSSL 296
QY 241 VSSIANNHMTSGVSMNTLSMLKGFAPAAAAQVQTAAGNGVAMSS----LGSSL 299
QY 297 GSSGLGGVAANLGRAASVGLSVPPQAAANQAVTPPAARALPLTSLTSAERGPQMLG 356
QY 300 GSSGLGAGVAANLGRAASVGLSVPPAWAANQAVTPPAARALPLTSLTSAQTAPGHMLG 359
QY 357 GLPVQMGARAGGGLSGVLRVPPRPVYMPHSPAAG 391
QY 360 GLPLGH-SVNAAGSGINNLRVPAAYAIPTPAAG 393

RESULT 7
Q7TJ3 PRELIMINARY; PRT; 393 AA.
AC Q7TJ3
DT 01-OCT-2003 (TReMBLrel. 25, Created)

```

```

DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE PPE family protein.
GN PPE26 OR MB1817.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97; PubMed=12788972;
RA Garnier T., Eigleier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248340; CAD94520.1; -.
KW Complete proteome.
SQ SEQUENCE 393 AA; 38572 MW; 4194436E547A8AB0 CRC64;

Query Match 40.0%; Score 779.5; DB 16; Length 393;
Best Local Similarity 43.7%; Pred. No. 2e-36;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 2 VDFGALPPEINSARMYAGPSASLVAAQWDSVADLFSASAFQSVVWGLTVGSWIGS 61
Db 1 MDFGALPPEINSARMYAGPSASLVAAQWDSVADLFSASAFQSVVWGLTVGSWIGS 60
QY 62 SAGLMVAASAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 121
Db 61 ASAAAEAVAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 122 IATNLLGQNTPAIAVNEAYEGEMWADAAAFYAAATATATATLPPPEAPEMTSAGGL 181
Db 121 ISTNVFGQNTSAIAAAEAQYEGEMWADAAAFYAAATATATATLPPPEAPEMTSAGGL 179
QY 182 LEOAAVEEASDIAIDTAANQLMNNVPOALQQLAQPAQWVPSKLGGLWTK----- 229
Db 180 GTQAAVAATAAGTAAGTATLTEMITGLPNALQSLTSLQSS-NGPLSLWQLLFGTNPFT 238
QY 230 -----TVSPHRSPISNMVSMANNHMTSGVSMNTLSMLKGFAPAAAAQVQTAAG 282
Db 239 SISALLTLQPYASFPYNTGLPYFISIGMNNFIOAKTL-GLIGSAAPAAVA-----AA 292
QY 283 QNGVRAMSLGSSGLSGGLGGVAANLGRAASVGLSVPOAWAA-ANOAVTPAARALPLT 341
Db 293 GDAAKGLPOLGMLG-----GGPVAAGLGNAAASVGLSVPPVNSGGLPGSVTFGAAPLVS 348
QY 342 SLTSAERGPQMLGGLPVQMGARAGGGLSGVLRVPPRPVYMPHSPAAG 391
Db 349 TVSAAPAEAPAGSLGGLPL----ACAGAGAGAP-RYGRPTVMARPPFAG 393

RESULT 8
OS3939 PRELIMINARY; PRT; 411 AA.
AC OS3939
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-VAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE PPE-family protein.
GN RV1789 OR MT1838 OR MTV049.11.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37FV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

```

RA Gordon S.V., Biglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horneby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence";
RL Nature 393:537-544 (1998).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RC Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022021; CAA17711.1; ALT_INIT.
DR EMBL; AB007043; AAK46108.1; -;
DR F1R; G70929; G70929.
DR TIGR; MT1838; -;
DR Tuberculist; RV1789; -;
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Complete proteome.
SQ SEQUENCE 411 AA; 40557 MW; 03C90B5E0590B7DA CRC64;

Query Match 40.0%; Score 779.5; DB 16; Length 411;
Best Local Similarity 43.7%; Pred. No. 2.1e-36;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 2 VDFGALPPEINSMYAGPGSALVAAQWDSVADLPSAASAFOSVWGLTVGWSIGS 61
DB 19 MDFGALPPEINSMYAGPGSALVAAQWDSVADLPSAASAFOSVWGLTVGWSIGS 78
QY 62 SAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAYETAYGLTVPPVIAENRAELMI 121
DB 79 ASAAAEAVAPYVAMSVTAGQAEHTAAQVRVAAAYETAYGLTVPPVIAENRAELMI 138
QY 122 IATNLLGONTPTAVNAEYGEYMAQDAAMFGYAAATATATATLTPFEAPEMTSAGL 181
DB 139 ISTNVFGONTPTAVNAEYGEYMAQDAAMFGYAAATATATATLTPFEAPEMTSAGL 197
QY 182 LEQAAVEEASDTAAA--NOLMNVPOALQQLAQP--TGTTPSSKLGSLWK-- 229
DB 198 GTQAAVATAGTAQSTLIEMITGLPNALQSLTSPQLQSS--NGPLSLWQLPFTENFT 256
QY 230 -----TVSPHRSPTSNMVMNHNHMTNSGVSMTNTLSMLKGFAPAAQAQVOTRA 282
DB 257 SISALLTDLPYASFFYNTEGLPYFTSGMGNFIQSAKTL--GLIGSAAPAAVA-----AA 310
QY 283 QNGVRAMSSIGSSIGSSGLGGVAAANIGRAASVGSLSVPQAWAA--ANOAVTPAARALPT 341
DB 311 GDAKGLPGJGGMLG-----GGPVAAGLGNAAVSKLSVPPVWNGPLPGSVTPGAAPLPVS 366
QY 342 SLTSAAREGPGQMLGGLPVGQMGARAGGSLGVLVRPPRPYVMPHSPAAG 391
DB 367 TVSAAPAEAFGLSLGLPLPL-----AGAGGAGAGP--RYGFRPTVMARPPFAG 411

RESULT 9
Q99Q11 ID Q99Q11 PRELIMINARY; PRT; 410 AA.
AC Q99Q11; (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE 01-OCI-2002 (Tremblrel. 22, Last annotation update)
GN MYC1808 OR OV1808.

OS Mycobacterium microti.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1806;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MYC 94-2272, and OV254;
RA Sivadon V., Heym B., Mazancourt P., Gaillard J.-L.;
RT "PPE RV1808 orthologue of Mycobacterium microti";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF335180; AAK20894.1; -;
DR EMBL; AF335179; AAK20893.1; -;
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
SQ SEQUENCE 410 AA; 40044 MW; 3B9DA3174655A5EA CRC64;

Query Match 38.7%; Score 754; DB 2; Length 410;
Best Local Similarity 44.5%; Pred. No. 5.7e-35;
Matches 177; Conservative 51; Mismatches 140; Indels 30; Gaps 10;

QY 1 MVDFGALPPEINSMYAGPGSALVAAQWDSVADLPSAASAFOSVWGLTVGWSIG 60
DB 1 MLDGALPPEINSMYAGPGSALVAAQWDSVADLPSAASAFOSVWGLTVGWSIG 60
QY 61 SSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 PSSITMAAAVAPYVAMSVTAGQAEHTAAQVRVAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGONTPTAVNAEYGEYMAQDAAMFGYAAATATATATLTPFEAPEMTSAG 180
DB 121 LVATNIFGONTPTAVNAEYGEYMAQDAAMFGYAAATATATATLTPFEAPEMTSAG 179
QY 181 LLEQAAVEEASDTAAA-----NOLMNVPOALQQLAQP--TGTTPSSKLGSLWKTVSP 233
DB 180 TAAQSAVVAAQAAAGAAASDDITAAQ--SQLISLLPSTLQSLA--TTATATASAG--WDTV-- 233
QY 234 HRSPTSNMVMNHNHMTNSGVSMTNTLSMLKGFAPAAQAQVOTAAQNGVRAMS 290
DB 234 -LQSTITLANTGLTPYSIIIGATPGGHWLFGQIL--GLAQNAPGVALLGPKAAAGALS 291
QY 291 SLGSSLSG-----SGLGGVAAANIGRAASVGSLSVPQAWAAANQAVTPAARALPTSLTS 345
DB 292 PLALPLRGYIADITPLGGGATGGIARAIYVGSLSVPQAWAAAPVWRAVASVLPQTGAAP 351
QY 346 A-AERGPGQMLGGLPVGQMGARAGGSL-----SGVLRY 377
DB 352 ALAAEPALGFQEMALSSLAGLAGTAGTVRSGAGARV 389

RESULT 10
O53957 ID O53957 PRELIMINARY; PRT; 409 AA.
AC O53957; (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DE 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE PPE-family protein.
GN RV1808 OR MT1856.1 OR MTV049.30.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horneby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,

Best Local Similarity 42.8%; Pred. No. 3.2e-34; Matches 182; Conservative 60; Mismatches 122; Indels 61; Gaps 13;

QY 2 VDFGALPPPEINSARMYAGPGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGWSIGS 61
 Db 1 MDPATLPPPEINSARMYAGPGSAPMLAAASAWHGLSABLSALSYSLTLTGEWHGP 60

QY 62 SAGLVAAASPYVAMSVTGAQAEHTAAQVRVAAAAYETAYGLTVPPPIAENRAELMIL 121
 Db 61 ASASMTAAAPYVAMSVTAVRAEQAGAAQAAAEAAFAATVPPPIAENRAELMIL 120

QY 122 IATNLLGONTFAIVNAEYGEWMAQDAAMFGYAAATATATATLLPPEAPEMTSAGGL 181
 Db 121 IATNVLGONAPAIATEAQYAEWMSQDAWAMGYAGASAAAT-QLTPPTPEVQTTNASGL 179

QY 182 LEQAAAVEEASDTAAA-----NQLMNVVPOALQQLAQTQGT-----PSKLG----- 225
 Db 180 AAGSAAIAHATGASAGAQTTLSQLIAAPSVLQGLSSSTAATSAGPSGLLIGSGSS 239

QY 226 ---GLWKTVPSPHRSPISNMVMANNMTNSGVSM-TNTLS---SMLKGFAPAAAQAV 278
 Db 240 WLDKWLALDPN-----SNFWNTIASSGLFLPSNTIAPFLGLLGGVAAAADAGDV 289

QY 279 QTAQNGVRAMSSLSGSSL-----GSSGLGGVAAANLGRAASVGSLSVPOQWAAAANQAVTP 333
 Db 290 LGATSG-----CLGALVAPLGSAGLGGTVAGLGNATVGLTSVPSWTAAPLASP 344

QY 334 AARAL---PLTSITSAAREGPGMLGGLPVQMGARAGGLSVLRVPP-----RPVMPH 386
 Db 345 LGSALGOTPMVAPPVAAAG---MFGMPFGTMGGQGF-----RAVPOYGRFNFVAR 394

QY 387 SPAAG 391
 Db 395 PPAAG 399

RESULT 13

O53956 PRELIMINARY; PRT; 403 AA.

ID O53956 (TrEMBLrel. 06, Created)
 AC O53956 (TrEMBLrel. 06, Last sequence update)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE PPE-family protein.
 GN RV1807 OR MT1656 OR MTV049, 29.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Ohkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and

laboratory strains";
 Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL022021; CAA17728.1; -;
 DR EMBL; AB007044; AAK46128.1; -;
 DR PIR; H70931; H70931.
 DR TIGR; MT1856; -;
 DR TubercuList; RV1807; -;
 DR InterPro; IPR000303; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Complete proteome. 227 F -> S (IN REF. 2).
 FT CONFLICT 238 V -> L (IN REF. 2).
 FT CONFLICT 238
 SQ SEQUENCE 403 AA; 39243 MW; DCE18880FD15CBFE CRC64;

Query Match 37.8%; Score 737.5; DB 16; Length 403;
 Best Local Similarity 42.6%; Pred. No. 4.8e-34;
 Matches 181; Conservative 60; Mismatches 123; Indels 61; Gaps 12;

QY 2 VDFGALPPPEINSARMYAGPGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGWSIGS 61
 Db 5 LDFATLPPPEINSARMYAGPGSAPMLAAASAWHGLSABLSALSYSLTLTGEWHGP 64

QY 62 SAGLVAAASPYVAMSVTGAQAEHTAAQVRVAAAAYETAYGLTVPPPIAENRAELMIL 121
 Db 65 ASASMTAAAPYVAMSVTAVRAEQAGAAQAAAEAAFAATVPPPIAENRAELMIL 124

QY 122 IATNLLGONTFAIVNAEYGEWMAQDAAMFGYAAATATATATLLPPEAPEMTSAGGL 181
 Db 125 IATNVLGONAPAIATEAQYAEWMSQDAWAMGYAGASAAAT-QLTPPTPEVQTTNASGL 183

QY 182 LEQAAAVEEASDTAAA-----NQLMNVVPOALQQLAQTQGT-----TPSS 222
 Db 184 AAGSAAIAHATGASAGAQTTLSQLIAAPSVLQGLSSSTAATSAGPSGLLIGSGSS 243

QY 223 KLGGLWKTVPSPHRSPISNMVMANNMTNSGVSM-TNTLS---SMLKGFAPAAAQAV 278
 Db 244 WLDKWLALDPN-----SNFWNTIASSGLFLPSNTIAPFLGLLGGVAAAADAGDV 293

QY 279 QTAQNGVRAMSSLSGSSL-----GSSGLGGVAAANLGRAASVGSLSVPOQWAAAANQAVTP 333
 Db 294 LGATSG-----CLGALVAPLGSAGLGGTVAGLGNATVGLTSVPSWTAAPLASP 348

QY 334 AARAL---PLTSITSAAREGPGMLGGLPVQMGARAGGLSVLRVPP-----RPVMPH 386
 Db 349 LGSALGOTPMVAPPVAAAG---MFGMPFGTMGGQGF-----RAVPOYGRFNFVAR 398

QY 387 SPAAG 391
 Db 399 PPAAG 403

RESULT 14

O53950 PRELIMINARY; PRT; 423 AA.

ID O53950 (TrEMBLrel. 06, Created)
 AC O53950 (TrEMBLrel. 06, Last sequence update)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE PPE-family protein.
 GN RV1801 OR MTV049, 23 OR MT1850.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Ohkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:07:54 ; Search time 26.0232 Seconds
(without alignments)
1446.225 Million cell updates/sec

Title: US-09-886-349A-16

Perfect score: 3686
Sequence: 1 MHHHHHTAASDNFQLSQGG.....SGGPVNLGQVGMNTAAS 729

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/prodata/2/iaa/PCOTUS_COMB.pep.*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3686	100.0	729	4	US-09-223-040-2
2	3686	100.0	729	4	US-09-287-849-2
3	2941	79.8	596	4	US-09-287-849-26
4	1967.5	53.4	600	4	US-09-287-849-22
5	1944	52.7	391	3	US-08-818-112-107
6	1944	52.7	391	4	US-08-818-111-102
7	1944	52.7	391	4	US-09-056-556-107
8	1944	52.7	391	4	US-09-072-967-102
9	1944	52.7	391	4	US-09-072-967-107
10	1652.5	44.8	396	3	US-08-818-112-111
11	1652.5	44.8	396	4	US-08-818-111-106
12	1652.5	44.8	396	4	US-09-056-556-111
13	1652.5	44.8	396	4	US-09-072-967-106
14	1652.5	44.8	396	4	US-09-072-967-111
15	1486.5	40.3	359	3	US-08-818-112-109
16	1486.5	40.3	359	4	US-08-818-111-104
17	1486.5	40.3	359	4	US-09-056-556-109
18	1486.5	40.3	359	4	US-09-072-967-104
19	1486.5	40.3	359	4	US-09-072-967-109
20	1184	32.1	263	4	US-09-287-849-8
21	1182	32.1	263	3	US-08-818-112-91
22	1182	32.1	263	4	US-08-818-111-92
23	1182	32.1	263	4	US-09-056-556-91
24	1182	32.1	263	4	US-09-072-967-92
25	1182	32.1	263	4	US-09-072-967-91
26	983	26.7	355	3	US-08-818-112-79
27	983	26.7	355	4	US-08-818-111-80

28	983	26.7	355	4	US-09-056-556-79	Sequence 79, Appl
29	983	26.7	355	4	US-09-072-967-80	Sequence 80, Appl
30	983	26.7	355	4	US-09-072-967-79	Sequence 79, Appl
31	766.5	20.8	400	4	US-09-073-009-126	Sequence 126, App
32	766.5	20.8	400	4	US-09-073-010-126	Sequence 126, App
33	743.5	20.2	231	4	US-09-287-849-28	Sequence 28, Appl
34	743.5	20.2	654	4	US-09-620-412C-341	Sequence 341, App
35	743.5	20.2	654	4	US-09-598-419-341	Sequence 341, App
36	741	20.1	518	4	US-09-620-412C-333	Sequence 333, App
37	741	20.1	518	4	US-09-598-419-333	Sequence 333, App
38	734	19.9	715	4	US-09-620-412C-321	Sequence 321, App
39	734	19.9	715	4	US-09-598-419-321	Sequence 321, App
40	733.5	19.9	715	4	US-09-620-412C-329	Sequence 329, App
41	733.5	19.9	715	4	US-09-598-419-329	Sequence 329, App
42	727	19.7	631	4	US-09-620-412C-325	Sequence 325, App
43	727	19.7	631	4	US-09-598-419-325	Sequence 325, App
44	723.5	19.6	525	4	US-09-556-877-196	Sequence 196, App
45	723.5	19.6	525	4	US-09-620-412C-196	Sequence 196, App

ALIGNMENTS

RESULT 1
US-09-223-040-2
; Sequence 2, Application US/09223040
; Patent No. 6544522
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-223-040-2

Query Match	100.0%;	Score 3686;	DB 4;	Length 729;
Best Local Similarity	100.0%;	Pred. No. 3.6e-261;		
Matches 729;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MHHHHHTAASDNFQLSQGGGFAIPICQAWAIAQIRSGGSPVTHIGPTAFGLGVVD	60	
DB	1	MHHHHHTAASDNFQLSQGGGFAIPICQAWAIAQIRSGGSPVTHIGPTAFGLGVVD	60	
QY	61	NNGGARVQVVGSAPAAASLGISTGDVITAVDGPINSATAMADALNHHGHPGDVISTWQ	120	
DB	61	NNGGARVQVVGSAPAAASLGISTGDVITAVDGPINSATAMADALNHHGHPGDVISTWQ	120	
QY	121	TKSGGTRTGNVTLAEGPPAEPMVDFGALPPEINARMYAGPGSASLVAAAQWDSVASDL	180	
DB	121	TKSGGTRTGNVTLAEGPPAEPMVDFGALPPEINARMYAGPGSASLVAAAQWDSVASDL	180	
QY	181	PSAASAFOSVVMGLTVGWSIGSSAGLMVAASPVVAMSVTAGOALTAQVRVAAAAYE	240	
DB	181	PSAASAFOSVVMGLTVGWSIGSSAGLMVAASPVVAMSVTAGOALTAQVRVAAAAYE	240	
QY	241	TAYGLTVPVPPVIAENRAELMILIAITNLGQNTPAIVNEAFYGEWMAQDAAMFGYAAAT	300	
DB	241	TAYGLTVPVPPVIAENRAELMILIAITNLGQNTPAIVNEAFYGEWMAQDAAMFGYAAAT	300	
QY	301	ATATATLLPFEPAPMTSAGGLFQAARVEASTAAANOLMNVVPAQLOLAQPTCGTT	360	

; PRIOR APPLICATION NUMBER: US 09/056,556
; CURRENT FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 79.8%; Score 2941; DB 4; Length 596;
Best Local Similarity 99.8%; Pred. No. 7.6e-207;
Matches 587; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 142 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201
Db 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68

QY 202 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261
Db 69 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 128

QY 262 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 321
Db 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 188

QY 322 LLEQAAAVEEASDTAAANQNMNVPOALQOLAQTOGTTPSSKLGGLWKTVPSPHSPI SN 381
Db 189 LLEQAAAVEEASDTAAANQNMNVPOALQOLAQTOGTTPSSKLGGLWKTVPSPHSPI SN 248

QY 382 MYSVANNHSMNTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 441
Db 249 MYSVANNHSMNTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 308

QY 442 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 501
Db 309 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 368

QY 502 GQMGARAGGGLSGVLVPRPYVMPHSPAAGDIAPPALSDQRPADFPALPLDPSAMVAQV 561
Db 369 GQMGARAGGGLSGVLVPRPYVMPHSPAAGDIAPPALSDQRPADFPALPLDPSAMVAQV 428

QY 562 GPQVWNTKLGYNNAVAGTGVIDPQGVVLTNNHVIAGTDINAFSGSGGTGYGVVY 621
Db 429 GPQVWNTKLGYNNAVAGTGVIDPQGVVLTNNHVIAGTDINAFSGSGGTGYGVVY 488

QY 622 GYDRTQDVAVLQLRGAGGLPSAAGGVAVGEPVWAMGNSGGGTTPRAVGRVVALGQT 681
Db 489 GYDRTQDVAVLQLRGAGGLPSAAGGVAVGEPVWAMGNSGGGTTPRAVGRVVALGQT 548

QY 682 VQASDSLTAETLNGLIQFDAAIQGDSGGPVVNGLGQVVGNTAAS 729
Db 549 VQASDSLTAETLNGLIQFDAAIQGDSGGPVVNGLGQVVGNTAAS 596

RESULT 4
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US

; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

Query Match 53.4%; Score 1967.5; DB 4; Length 600;
Best Local Similarity 72.0%; Pred. No. 1e-135;
Matches 431; Conservative 19; Mismatches 82; Indels 67; Gaps 10;

QY 142 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201
Db 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68

QY 202 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261
Db 69 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 128

QY 262 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 321
Db 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 188

QY 322 LLEQAAAVEEASDTAAANQNMNVPOALQOLAQTOGTTPSSKLGGLWKTVPSPHSPI SN 381
Db 189 LLEQAAAVEEASDTAAANQNMNVPOALQOLAQTOGTTPSSKLGGLWKTVPSPHSPI SN 248

QY 382 MYSVANNHSMNTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 441
Db 249 MYSVANNHSMNTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 308

QY 442 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 501
Db 309 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABRGPGQMLGGLPV 368

QY 502 GQMGARAGGGLSGVLVPRPYVMPHSPAAGDIAPPALSDQRPADFPALPLDPSAMVAQV 543
Db 369 GQMGARAGGGLSGVLVPRPYVMPHSPAAGDIAPPALSDQRPADFPALPLDPSAMVAQV 428

QY 544 FADFPALPLDPS-----AMVAQVGPQVWNTKLGYNNAVAGTGVIDPN 589
Db 429 AAQFNASPVASQSYLRNFLLAAPPPORAAQAOL-QAVPGAAQVIGLVESVAGSCN-----N 482

QY 590 GVVLTTNNHVIAGTDINAFSGSGGTGYGVVLTNNHVIAGTDINAFSGSGGTGYGVVY 649
Db 483 YELMTINQCFG---DVDA---HGAMIRAQAASLEAEHQAIVRDVLAAGDFWGA--GSV 533

QY 650 AVGEPVWAMG-----NSGGGGGTTPRAVGRVVALGQTVQASDSLTAETLNG 697
Db 534 ACQEPITQIGRNFQVYEQANAHGQ-----KVQAAGNNMAQTDSAVGSSWATSNG 583

RESULT 5
US-08-818-112-107
; Sequence 107, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

Qy 502 QMGARAGGGGSLGVLRPPRPVYMPHSPAAG 532
Dd 361 QMGARAGGGGSLGVLRPPRPVYMPHSPAAG 391

RESULT 7

```

US-09-056-556-107
/ Sequence 107, Application US/09056556
/ Patent No. 6350456
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Dillon, Davin C.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
/ NUMBER OF SEQUENCES: 241
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SEED and BERRY LLP
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: USA
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/056,556
/ FILING DATE: 07-APR-1998
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Maki, David J.
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.457
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 107:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 391 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
US-09-056-556-107

```

Query Match	52.7%;	Score 1944;	DB 4;	Length 391;
Best Local Similarity	99.7%;	Pred. No. 3e-134;		
Matches 390;	Conservative	0;	Mismatches 1;	Indels 0;
Gaps	0;			

[illegible]

Qy 502 QQMGARAGGGTSGVLRVPPRPYVNEPHSPAAG 532
|||
Db 361 QQMGARAGGGTSGVLRVPPRPYVNEPHSPAAG 391

RESULT 8

```

US-09-072-596-102
; Sequence 102 Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-102

```

Query Match 52.7%; Score 1944; DB 4; Length 391;
Best Local Similarity 99.7%; Pred. No. 3e-134;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

142	QY	MVDFGALPPEINSARMYAGPGSASLVAQAQMDMSVASDLFSAASAFQSVVWGLTVGSMWG	201
1	Db	MVDFGALPPEINSARMYAGPGSASLVAQAQMDMSVASDLFSAASAFQSVVWGLTVGSMWG	60
202	QY	SSAGLMTVAASPYVAMVSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPVPVIAENAEALMI	261
61	Db	SSAGLMTVAASPYVAMVSVTAGQAEITAAQVRVAAAAAYETAYGLTVPPVPVIAENAEALMI	120
262	QY	LIATNLGQNTPTAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEEAPEMTSAGG	321
121	Db	LIATNLGQNTPTAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEEAPEMTSAGG	180
322	QY	LLEQAAAVEBASDTAAANLMMNVPAQLQALQOQTGCTTFSKGLGGLWKTVSPHSPISN	381
181	Db	LLEQAAAVEBASDTAAANO.MNNVPAQLQALQOQTGCTTFSKGLGGLWKTVSPHSPISN	240

QY 382 MVSMMNHMTNSGVSMNTLSSMLKGFAPAAARQAVQTAQAQNGVRAMSSLSLSSGSSG 441
DB 241 MVSMMNHMTNSGVSMNTLSSMLKGFAPAAARQAVQTAQAQNGVRAMSSLSLSSGSSG 300
QY 442 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERPGQMLGGLPV 501
DB 301 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERPGQMLGGLPV 360
QY 502 GOMGARAGGSLGVLRVPPRPVMPHSPAAG 532
DB 361 GOMGARAGGSLGVLRVPPRPVMPHSPAAG 391

RESULT 9
US-09-072-967-107
; Sequence 107, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-09-072-967-107

Query Match 52.7%; Score 1944; DB 4; Length 391;
Best Local Similarity 99.7%; Pred. No. 3e-134;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 142 MVDFGALPPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 201
DB 1 MVDFGALPPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY 202 SSAGLMVAASPPYVWMSVTAQAEILTAQVRVAAAAYETAYGLTVPVIAENRAELMI 261
DB 61 SSAGLMVAASPPYVWMSVTAQAEILTAQVRVAAAAYETAYGLTVPVIAENRAELMI 120

QY 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLILPFEPAPMTSAGG 321
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLILPFEPAPMTSAGG 180
QY 322 LLEQAAVEEASDTPAANQLMNNVPOALQOAOPTQGTTPSSKLGGLWKVSPHRSPLSN 381
DB 181 LLEQAAVEEASDTPAANQLMNNVPOALQOAOPTQGTTPSSKLGGLWKVSPHRSPLSN 240
QY 382 MVSMMNHMTNSGVSMNTLSSMLKGFAPAAARQAVQTAQAQNGVRAMSSLSLSSGSSG 441
DB 241 MVSMMNHMTNSGVSMNTLSSMLKGFAPAAARQAVQTAQAQNGVRAMSSLSLSSGSSG 300
QY 442 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERPGQMLGGLPV 501
DB 301 LGGVVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERPGQMLGGLPV 360
QY 502 GOMGARAGGSLGVLRVPPRPVMPHSPAAG 532
DB 361 GOMGARAGGSLGVLRVPPRPVMPHSPAAG 391

RESULT 10
US-08-818-112-111
; Sequence 111, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

US-08-818-112-111

Query Match 44.8%; Score 1652.5; DB 3; Length 396;
Best Local Similarity 84.9%; Pred. No. 6.2e-113;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 142 MVDFGALPPPEINARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 201

```

Db 1 VDFGALPEINSAARMYAGPGSASLVAAAKWMDSVASDLFSAASAFQSVVWGLTGSWIG 60
QY 202 SSAGLMVAAASPYPVAMNSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 261
Db 61 SSAGLMVAAASPYPVAMNSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 120
QY 262 LIATNLLGONTPTAIAVNEAEYGEWMAQDAAMFGYAAATATATATATALLPFEAPEMTSAGG 321
Db 121 LIATNLLGONTPTAIAVNEAEYGEWMAQDAAMFGYAAATATATATATALLPFEAPEMTSAGG 180
QY 322 LLEQAAAVEEASDTAAANQLMNNVPQALQOQAQPTQTSKIGGLWKTVPSPHRSPI SN 381
Db 181 LLEQAAVEEAIDTAAANQLMNNVPQALQOQAQPTQTSKINWPDQLSELWKAI SPHLSPLSN 240
QY 382 MVSAMNNHMTNSGVSMTNTLSMKLGFAPAAARQAVTAAQNGVRAMSS----LGSSL 437
Db 241 IVSMNNHVSMTNSGVSMTNTLSMKLGFAPAAA-QAVETAQNGVQAMSSLGSSLGSSL 299
QY 438 GSSGLGGVAANLGRAASVGSLSVPQAWAAANQAVTAAARALPLTSLTSAABERGPGOMLG 497
Db 300 GSSGLGAGVAANLGRAASVGSLSVPQAWAAANQAVTAAARALPLTSLTSAABERGPGOMLG 359

RESULT 11
US-08-818-111-106
; Sequence 106, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-818-111-106

```

Query Match 44.8%; Score 1652.5; DB 4; Length 396;

```

Best Local Similarity 84.9%; Pred. No. 6.2e-113;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 142 MVDFGALPEINSAARMYAGPGSASLVAAAKWMDSVASDLFSAASAFQSVVWGLTGSWIG 201
Db 1 VDFGALPEINSAARMYAGPGSASLVAAAKWMDSVASDLFSAASAFQSVVWGLTGSWIG 60
QY 202 SSAGLMVAAASPYPVAMNSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 261
Db 61 SSAGLMVAAASPYPVAMNSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIVIAENRAELMI 120
QY 262 LIATNLLGONTPTAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPEMTSAGG 321
Db 121 LIATNLLGONTPTAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPEMTSAGG 180
QY 322 LLEQAAAVEEASDTAAANQLMNNVPQALQOQAQPTQTSKIGGLWKTVPSPHRSPI SN 381
Db 181 LLEQAAVEEAIDTAAANQLMNNVPQALQOQAQPTQTSKINWPDQLSELWKAI SPHLSPLSN 240
QY 382 MVSAMNNHMTNSGVSMTNTLSMKLGFAPAAARQAVTAAQNGVRAMSS----LGSSL 437
Db 241 IVSMNNHVSMTNSGVSMTNTLSMKLGFAPAAA-QAVETAQNGVQAMSSLGSSLGSSL 299
QY 438 GSSGLGGVAANLGRAASVGSLSVPQAWAAANQAVTAAARALPLTSLTSAABERGPGOMLG 497
Db 300 GSSGLGAGVAANLGRAASVGSLSVPQAWAAANQAVTAAARALPLTSLTSAABERGPGOMLG 359

RESULT 12
US-09-056-556-111
; Sequence 111, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-056-556-111

```

Query Match 44.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 6.2e-113;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-596-106

142 MVDFGALPPEINARMYAGPGSASLVAAQWSDVSAFQSVVWGLTVGSGWIG 201
Db :|||||
1 VVDFGALPPEINARMYAGPGSASLVAAQWSDVSAFQSVVWGLTVGSGWIG 60
Qy :|||||
202 SSAGLMVAAAAPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
Qy :|||||
61 SSAGLMVAAAAPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db :|||||
262 LIATNLLGONTPTAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 321
Qy :|||||
121 LIATNLLGONTPTAVNEAEYGEWMAQDAAMFGYAAATATATLPPFEAPMTSAGG 180
Db :|||||
322 LLEQAAVEEAASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 381
Qy :|||||
181 LLEQAAVEEAASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
Db :|||||
382 MYSMANNHMTNSGVSMTNTLSMLKGFAPAAAQVTAARALPLTSLTSAARPGQMLG 437
Qy :|||||
241 IVSMLNHNHVSMTNSGVSMASTLHSLKGFAPAAA-QAVETAQNGVQMSLSGLGSSL 299
Db :|||||
438 GSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAARPGQMLG 497
Qy :|||||
300 GSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAARPGHMLG 359
Db :|||||

RESULT 13

US-09-072-596-106
; Sequence 106, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072.596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:

Query Match 44.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 6.2e-113;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-596-106

142 MVDFGALPPEINARMYAGPGSASLVAAQWSDVSAFQSVVWGLTVGSGWIG 201
Db :|||||
1 VVDFGALPPEINARMYAGPGSASLVAAQWSDVSAFQSVVWGLTVGSGWIG 60
Qy :|||||
202 SSAGLMVAAAAPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
Qy :|||||
61 SSAGLMVAAAAPYVAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
Db :|||||
262 LIATNLLGONTPTAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 321
Qy :|||||
121 LIATNLLGONTPTAVNEAEYGEWMAQDAAMFGYAAATATATLPPFEAPMTSAGG 180
Db :|||||
322 LLEQAAVEEAASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 381
Qy :|||||
181 LLEQAAVEEAASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
Db :|||||
382 MYSMANNHMTNSGVSMTNTLSMLKGFAPAAAQVTAARALPLTSLTSAARPGQMLG 437
Qy :|||||
241 IVSMLNHNHVSMTNSGVSMASTLHSLKGFAPAAA-QAVETAQNGVQMSLSGLGSSL 299
Db :|||||
438 GSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAARPGQMLG 497
Qy :|||||
300 GSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAARPGHMLG 359
Db :|||||

RESULT 14

US-09-072-967-111
; Sequence 111, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072.967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:


```
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-09-072-967-111

Query Match 44.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 6.2e-113;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 142 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201
Db 1 VVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 202 SSAGLMVAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 261
Db 61 SSAGLMVAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
QY 262 LIATNLLGONTPTAIIVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 321
Db 121 LIATNLLGONTPTAIIVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180
QY 322 LLEQAAVVEEASDTAAANQLMNNVPOALQQAQPTOGTTPSSKLGGLWKTVPSPHSPLSN 381
Db 181 LLEQAVAVEEADTAAANQLMNNVPOALQQAQPTOGTTPSSKLGGLWKTVPSPHSPLSN 240
QY 382 MYSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMS-----LGSSL 437
Db 241 IVSMLNHYSMTNSGVSMNTLSSMLKGFAPAAA-QAVETAQNGVQAQMSLSGSL 299
QY 438 GSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQWMLG 497
Db 300 GSSGLGAGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAQTAPGHMLG 359

RESULT 15
US-08-818-112-109
; Sequence 109, Application US/08818112
; Patent No. 6290369
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESS: SEED and BERRY LLP
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-818-112-109

Query Match 40.3%; Score 1486.5; DB 3; Length 359;
Best Local Similarity 84.2%; Pred. No. 7.3e-101;
Matches 303; Conservative 16; Mismatches 36; Indels 5; Gaps 2;

QY 142 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201
Db 1 VVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 202 SSAGLMVAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 261
Db 61 SSAGLMVAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
QY 262 LIATNLLGONTPTAIIVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 321
Db 121 LIATNLLGONTPTAIIVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEEAPMTSAGG 180
QY 322 LLEQAAVVEEASDTAAANQLMNNVPOALQQAQPTOGTTPSSKLGGLWKTVPSPHSPLSN 381
Db 181 LLEQAVAVEEADTAAANQLMNNVPOALQQAQPTOGTTPSSKLGGLWKTVPSPHSPLSN 240
QY 382 MYSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMS-----LGSSL 437
Db 241 VSSIANNHMSMTNSGVSMNTLSSMLKGFAPAAA-QAVETAQNGVQAQMSLSGSL 299
QY 438 GSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGPQWMLG 497
Db 300 GSSGLGAGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAQTAPGHMLG 359

Search completed: June 22, 2004, 17:27:34
Job time : 28.0232 secs
```



GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 67.5195 Seconds
(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-16
Perfect score: 3686
Sequence: 1 MHHHHHTAASDNFOLSQGG.....SCGFVNLGQVGVNMTAAS 729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282113646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3686	100.0	729	9	US-09-287-849-2
2	3686	100.0	729	12	US-09-886-349A-16
3	3686	100.0	729	14	US-10-359-460-2
4	3686	100.0	729	14	US-10-098-732A-16
5	3686	100.0	729	15	US-10-359-459-2
6	3680	99.8	729	15	US-10-369-983-21
7	3680	99.8	813	15	US-10-369-983-15
8	3680	99.8	825	15	US-10-369-983-14
9	3680	99.8	875	15	US-10-369-983-13
10	3680	99.8	930	14	US-10-098-732A-65
11	3680	99.8	930	15	US-10-369-983-12
12	3680	99.8	1016	15	US-10-369-983-18
13	3680	99.8	1022	15	US-10-369-983-17
14	3680	99.8	1154	15	US-10-369-983-16
15	3677	99.8	729	12	US-09-886-349A-18

16	3677	99.8	729	14	US-10-098-732A-18
17	3677	99.8	729	15	US-10-369-983-22
18	2941	79.8	596	9	US-09-287-849-26
19	2941	79.8	596	12	US-09-886-349A-20
20	2941	79.8	596	14	US-10-359-460-26
21	2941	79.8	596	14	US-10-098-732A-20
22	2637	71.5	1010	15	US-10-369-983-4
23	2631	71.5	723	15	US-10-369-983-2
24	1967.5	53.4	600	9	US-09-287-849-22
25	1967.5	53.4	600	14	US-10-359-460-22
26	1944	52.7	391	12	US-09-886-349A-14
27	1944	52.7	391	14	US-10-193-002-102
28	1944	52.7	391	14	US-10-084-843-107
29	1944	52.7	391	14	US-10-098-732A-14
30	1929	52.3	391	12	US-09-872-186-8
31	1652.5	44.8	391	14	US-10-193-002-106
32	1652.5	44.8	396	14	US-10-084-843-111
33	1583	42.9	393	12	US-10-282-122A-62455
34	1583	42.9	393	12	US-10-282-122A-64892
35	1486.5	40.3	359	14	US-10-193-002-104
36	1486.5	40.3	359	14	US-10-084-843-109
37	1184	32.1	358	9	US-09-287-849-8
38	1184	32.1	358	14	US-10-359-460-8
39	1182	32.1	263	12	US-09-886-349A-12
40	1182	32.1	263	14	US-10-193-002-92
41	1182	32.1	263	14	US-10-084-843-91
42	1182	32.1	263	14	US-10-098-732A-12
43	989	26.8	355	9	US-09-712-363-161
44	987	26.8	330	12	US-09-886-349A-4
45	987	26.8	330	14	US-10-098-732A-4

ALIGNMENTS

RESULT 1

US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; NUMBER OF SEQ IDS NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-2

Query Match 100.0%; Score 3686; DB 9; Length 729;
Best Local Similarity 100.0%; Pred. No. 7.4e-244;

```

Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFOLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFOLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
QY 61 NNGNGARVQVRVVGSAAPASLSGISTGCVITAVDGPINSATAMADALNGHHPGCDVISVTWQ 120
DB 61 NNGNGARVQVRVVGSAAPASLSGISTGCVITAVDGPINSATAMADALNGHHPGCDVISVTWQ 120
QY 121 TKSGGTRTGNVTLAGPPAEFWDGALPPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
DB 121 TKSGGTRTGNVTLAGPPAEFWDGALPPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWISSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSGWISSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVPIAENRAELMILIAITNLLGQNTPAIANNVNEAYEGEMWADAAAFGYAAAT 300
DB 241 TAYGLTVPPVPIAENRAELMILIAITNLLGQNTPAIANNVNEAYEGEMWADAAAFGYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGLLLEQAAAVEASDTAAANOLMNNVPOALQOLAQPTQGT 360
DB 301 ATATATLLPFEAPEMTSAGLLLEQAAAVEASDTAAANOLMNNVPOALQOLAQPTQGT 360
QY 361 PSSKLGGLWKTVPVSHRSPISNNVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTVPVSHRSPISNNVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLSGSLGGVAAANLGRASVGSLSVPOQAAANQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLSGSLGGVAAANLGRASVGSLSVPOQAAANQAVTPAARALP 480
QY 481 LTSLSAAERPGQMLGGLPVQMGARAGGSLGVLVPPRPVPMHSPSAAAGDIAPPALS 540
DB 481 LTSLSAAERPGQMLGGLPVQMGARAGGSLGVLVPPRPVPMHSPSAAAGDIAPPALS 540
QY 541 QDRFADFPALFDPDSAMVAQVGPQVNTKLGYNNAVGAGTGIVDPNGVLTNNHVA 600
DB 541 QDRFADFPALFDPDSAMVAQVGPQVNTKLGYNNAVGAGTGIVDPNGVLTNNHVA 600
QY 601 GATDINAFSVGSGQTYGVVDVYDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMAGN 660
DB 601 GATDINAFSVGSGQTYGVVDVYDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMAGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNLGLIOFDAAIQPGDSGGPVVNLGQ 720
DB 661 SGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNLGLIOFDAAIQPGDSGGPVVNLGQ 720
QY 721 VVGNNMTAAS 729
DB 721 VVGNNMTAAS 729

```

RESULT 2

```

US-09-886-349A-16
; Sequence 16, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-09070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50

```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein MTB72f (Ra12-TbH9-Ra35 or MTB32-MTB39
; OTHER INFORMATION: fusion)
US-09-886-349A-16

```

```

Query Match 100.0%; Score 3686; DB 12; Length 729;
Best Local Similarity 100.0%; Pred. No. 7.4e-244;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MHHHHHTAASDNFOLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFOLSGGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
QY 61 NNGNGARVQVRVVGSAAPASLSGISTGCVITAVDGPINSATAMADALNGHHPGCDVISVTWQ 120
DB 61 NNGNGARVQVRVVGSAAPASLSGISTGCVITAVDGPINSATAMADALNGHHPGCDVISVTWQ 120
QY 121 TKSGGTRTGNVTLAGPPAEFWDGALPPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
DB 121 TKSGGTRTGNVTLAGPPAEFWDGALPPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWISSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSGWISSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVPIAENRAELMILIAITNLLGQNTPAIANNVNEAYEGEMWADAAAFGYAAAT 300
DB 241 TAYGLTVPPVPIAENRAELMILIAITNLLGQNTPAIANNVNEAYEGEMWADAAAFGYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGLLLEQAAAVEASDTAAANOLMNNVPOALQOLAQPTQGT 360
DB 301 ATATATLLPFEAPEMTSAGLLLEQAAAVEASDTAAANOLMNNVPOALQOLAQPTQGT 360
QY 361 PSSKLGGLWKTVPVSHRSPISNNVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTVPVSHRSPISNNVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLSGSLGGVAAANLGRASVGSLSVPOQAAANQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLSGSLGGVAAANLGRASVGSLSVPOQAAANQAVTPAARALP 480
QY 481 LTSLSAAERPGQMLGGLPVQMGARAGGSLGVLVPPRPVPMHSPSAAAGDIAPPALS 540
DB 481 LTSLSAAERPGQMLGGLPVQMGARAGGSLGVLVPPRPVPMHSPSAAAGDIAPPALS 540
QY 541 QDRFADFPALFDPDSAMVAQVGPQVNTKLGYNNAVGAGTGIVDPNGVLTNNHVA 600
DB 541 QDRFADFPALFDPDSAMVAQVGPQVNTKLGYNNAVGAGTGIVDPNGVLTNNHVA 600
QY 601 GATDINAFSVGSGQTYGVVDVYDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMAGN 660
DB 601 GATDINAFSVGSGQTYGVVDVYDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMAGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNLGLIOFDAAIQPGDSGGPVVNLGQ 720
DB 661 SGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNLGLIOFDAAIQPGDSGGPVVNLGQ 720
QY 721 VVGNNMTAAS 729
DB 721 VVGNNMTAAS 729

```

RESULT 3

```

US-10-359-460-2
; Sequence 2, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:

```

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Algerson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/10/359,460
PRIOR FILING DATE: 2003-02-05
PRIOR APPLICATION NUMBER: US/09/287,849
PRIOR FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 2
LENGTH: 729
TYPE: PRT
ORGANISM: Artificial Sequence
OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-10-359-460-2

Query Match 100.0%; Score 3686; DB 14; Length 729;
Best Local Similarity 100.0%; Pred. No. 7,4e-244;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60

QY 61 NNGNGARVQVVGSAFAPASIGITGVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
Db 61 NNGNGARVQVVGSAFAPASIGITGVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120

QY 121 TKSGETRTGNVTLAEGPPAEFMDVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
Db 121 TKSGETRTGNVTLAEGPPAEFMDVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSGWISAGLMVAASPYVAMWSVTAGQAEITAAQVRAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWISAGLMVAASPYVAMWSVTAGQAEITAAQVRAAAAYE 240

QY 241 TAYGLTVPPVPIAENRAELMILITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVPIAENRAELMILITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300

QY 301 ATATATLLPFEAPEMTSAGLLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGLLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGT 360

QY 361 PSSKLGGLWKTVPSPHRSPISNNVSMNNHSMNTSGVSMNTLSSMLKGFAPAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNNVSMNNHSMNTSGVSMNTLSSMLKGFAPAAARQAVQ 420

QY 421 TAAQNGVRAMSSLSGSSGLGGVAANTGRAASVGSLSVPQAWAANAQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLSGSSGLGGVAANTGRAASVGSLSVPQAWAANAQAVTPAARALP 480

QY 481 LTSLSAERPGQMLGLPVGOMGARAGGLSGVLVPRPPYVPHSPAGDIAPFALS 540
Db 481 LTSLSAERPGQMLGLPVGOMGARAGGLSGVLVPRPPYVPHSPAGDIAPFALS 540

QY 541 QDRFADFPALPLDPSAMVAQVGFQVNVNITKLGNNAVAGTGTIVDPNGVLTNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGFQVNVNITKLGNNAVAGTGTIVDPNGVLTNNHVA 600

QY 601 GATDINAFSVSGGTGYGVVDVAVLQLRGAGGLPSAAIGGVAAGVGPVVMAGN 660
Db 601 GATDINAFSVSGGTGYGVVDVAVLQLRGAGGLPSAAIGGVAAGVGPVVMAGN 660

QY 661 SGGGGTFRAPVGRVVALGQTVQASDLTGAEETLNGLIQFDAAIQPDGSGGPVNLGQ 720
Db 661 SGGGGTFRAPVGRVVALGQTVQASDLTGAEETLNGLIQFDAAIQPDGSGGPVNLGQ 720

QY 721 VVGNMTAAS 729
Db 721 VVGNMTAAS 729

RESULT 4
US-10-098-732A-16
; Sequence 16, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 16
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein MTB7F (Ra12-1bH9-Ra35 or MTB32-MTB39
; OTHER INFORMATION: fusion)
US-10-098-732A-16

Query Match 100.0%; Score 3686; DB 14; Length 729;
Best Local Similarity 100.0%; Pred. No. 7,4e-244;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQSQGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60

QY 61 NNGNGARVQVVGSAFAPASIGITGVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
Db 61 NNGNGARVQVVGSAFAPASIGITGVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120

QY 121 TKSGETRTGNVTLAEGPPAEFMDVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
Db 121 TKSGETRTGNVTLAEGPPAEFMDVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSGWISAGLMVAASPYVAMWSVTAGQAEITAAQVRAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWISAGLMVAASPYVAMWSVTAGQAEITAAQVRAAAAYE 240

QY 241 TAYGLTVPPVPIAENRAELMILITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVPIAENRAELMILITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300

QY 301 ATATATLLPFEAPEMTSAGLLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGLLEQAAAVEASDTAAANQLMNNVPQALQQAQPTQGT 360

QY 361 PSSKLGGLWKTVSPHRSPIISNNVSMANNHMTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
 DB 361 PSSKLGGLWKTVSPHRSPIISNNVSMANNHMTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
 QY 421 TAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTAAARALP 480
 DB 421 TAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTAAARALP 480
 QY 481 LTSLSAAERGPQMLGGLPVQMGARAGGSLGVLRVPPRYPMHSPHSPAGDIAPPALS 540
 DB 481 LTSLSAAERGPQMLGGLPVQMGARAGGSLGVLRVPPRYPMHSPHSPAGDIAPPALS 540
 QY 541 QDRFADFPALPLDPSAMVAQVGPQVNTNKLGYNNVAGAGTGIVDPNGVVLNNHVA 600
 DB 541 QDRFADFPALPLDPSAMVAQVGPQVNTNKLGYNNVAGAGTGIVDPNGVVLNNHVA 600
 QY 601 GATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPPVAMGN 660
 DB 601 GATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPPVAMGN 660
 QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDSGGPVVNLGQ 720
 DB 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDSGGPVVNLGQ 720
 QY 721 VVGMMNTAAS 729
 DB 721 VVGMMNTAAS 729

RESULT 5
 US-10-359-459-2
 ; Sequence 2, Application US/10359459
 ; Publication No. US20040013677A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
 ; TITLE OF INVENTION: and Their Uses
 ; FILE REFERENCE: 014058-009010US
 ; CURRENT APPLICATION NUMBER: US/10/359,459
 ; CURRENT FILING DATE: 2003-02-05
 ; PRIOR APPLICATION NUMBER: US/09/223,040
 ; PRIOR FILING DATE: 1998-12-30
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 729
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
 US-10-359-459-2

Query Match 100.0%; Score 3686; DB 15; Length 729;
 Best Local Similarity 100.0%; Pred. No. 7.4e-244;
 Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPVTHIGPTAFGLGVVD 60
 DB 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPVTHIGPTAFGLGVVD 60
 QY 61 NNGNGARVQVVGSAAPASLIGSTGDIITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
 DB 61 NNGNGARVQVVGSAAPASLIGSTGDIITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
 QY 121 TKSGGTTRTNVLAEGPPAEFMYDFCALPEINSARWYAGPSASLVAAQWDSVSDL 180
 DB 121 TKSGGTTRTNVLAEGPPAEFMYDFCALPEINSARWYAGPSASLVAAQWDSVSDL 180
 QY 181 FSAASAFQSVWGLTVGSGWIGSAGLMVAASPPYVAMSVTAQAEHTAAQVRVAAAAYE 240

DB 181 FSAASAFQSVWGLTVGSGWIGSAGLMVAASPPYVAMSVTAQAEHTAAQVRVAAAAYE 240
 QY 241 TAYGLTVPPPVIAENRAELMILIAITNLGQNTPAIAVNEAEYGEWMAQDAARAFVAAAT 300
 DB 241 TAYGLTVPPPVIAENRAELMILIAITNLGQNTPAIAVNEAEYGEWMAQDAARAFVAAAT 300
 QY 301 ATATATLLPFEAPEMTSAGGLEQAAAEEASDSTAANQLMNNVFOALQQLAQPTQGT 360
 DB 301 ATATATLLPFEAPEMTSAGGLEQAAAEEASDSTAANQLMNNVFOALQQLAQPTQGT 360
 QY 361 PSSKLGGLWKTVSPHRSPIISNNVSMANNHMTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
 DB 361 PSSKLGGLWKTVSPHRSPIISNNVSMANNHMTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
 QY 421 TAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTAAARALP 480
 DB 421 TAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTAAARALP 480
 QY 481 LTSLSAAERGPQMLGGLPVQMGARAGGSLGVLRVPPRYPMHSPHSPAGDIAPPALS 540
 DB 481 LTSLSAAERGPQMLGGLPVQMGARAGGSLGVLRVPPRYPMHSPHSPAGDIAPPALS 540
 QY 541 QDRFADFPALPLDPSAMVAQVGPQVNTNKLGYNNVAGAGTGIVDPNGVVLNNHVA 600
 DB 541 QDRFADFPALPLDPSAMVAQVGPQVNTNKLGYNNVAGAGTGIVDPNGVVLNNHVA 600
 QY 601 GATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPPVAMGN 660
 DB 601 GATDINAFSVSGGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPPVAMGN 660
 QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDSGGPVVNLGQ 720
 DB 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPDSGGPVVNLGQ 720
 QY 721 VVGMMNTAAS 729
 DB 721 VVGMMNTAAS 729

RESULT 6
 US-10-369-983-21
 ; Sequence 21, Application US/10369983
 ; Publication No. US20030235593A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Guderian, Jeff
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009081US
 ; CURRENT APPLICATION NUMBER: US/10/369,983
 ; CURRENT FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: US 60/357,351
 ; PRIOR FILING DATE: 2002-02-15
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 21
 ; LENGTH: 729
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: MTB72F
 US-10-369-983-21

Query Match 99.8%; Score 3680; DB 15; Length 729;
 Best Local Similarity 99.9%; Pred. No. 1.9e-243;
 Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPVTHIGPTAFGLGVVD 60
 DB 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPVTHIGPTAFGLGVVD 60
 QY 61 NNGNGARVQVVGSAAPASLIGSTGDIITAVDGPINSATAMADALNGHHPGDVSVTWQ 120

Db 61 NNGNGARVQVRVGSAPAAASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISVTWQ 120
QY 121 TKSGETRTGNVTLAEGPPAEFVDFGALPPEINSAARMYAGPGSASLVAAAQWDSVASDL 180
Db 121 TKSGETRTGNVTLAEGPPAEFVDFGALPPEINSAARMYAGPGSASLVAAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWIGSSAGLMVAASAPYVAMMSVTAGQAELTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWIGSSAGLMVAASAPYVAMMSVTAGQAELTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPPVIAENRAELMILIAITNLLGQNTPAIAVNEAEYGENWAOQAAAMFYAAAT 300
Db 241 TAYGLTVPPPVIAENRAELMILIAITNLLGQNTPAIAVNEAEYGENWAOQAAAMFYAAAT 300
QY 301 ATATATLLPFEAPPEMTSAGGLLEQAAAEEASDTAAANQLMNNVPOALQQLAQTQGT 360
Db 301 ATATATLLPFEAPPEMTSAGGLLEQAAAEEASDTAAANQLMNNVPOALQQLAQTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAARQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSSGGLGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSLGSSGGLGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
QY 481 LTSLSAAERPGQMLGGLPVGQMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
Db 481 LTSLSAAERPGQMLGGLPVGQMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPDPDSAMVAQVGPQVNVINTKLGYNNAVAGAGTGIVIDPENGVLTNHVTIA 600
Db 541 QDRFADFPALPDPDSAMVAQVGPQVNVINTKLGYNNAVAGAGTGIVIDPENGVLTNHVTIA 600
QY 601 GATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Db 601 GATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
QY 661 SGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
Db 661 SGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
QY 721 VVGMTAAS 729
Db 721 VVGMTAAS 729

RESULT 7
US-10-369-983-15
; Sequence 15, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB81F (MTB72F-DPV)
US-10-369-983-15

Query Match 99.8%; Score 3680; DB 15; Length 813;
Best Local Similarity 99.9%; Pred. No. 2.2e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MHHHHHHTAASDNFOLISOGGQGFALPIGQAMATAGQIRSGGSPTHVHGIFATFLGLGWD 60
Db 1 MHHHHHHTAASDNFOLISOGGQGFALPIGQAMATAGQIRSGGSPTHVHGIFATFLGLGWD 60
QY 61 NNGNGARVQVRVGSAPAAASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISVTWQ 120
Db 61 NNGNGARVQVRVGSAPAAASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISVTWQ 120
QY 121 TKSGETRTGNVTLAEGPPAEFVDFGALPPEINSAARMYAGPGSASLVAAAQWDSVASDL 180
Db 121 TKSGETRTGNVTLAEGPPAEFVDFGALPPEINSAARMYAGPGSASLVAAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWIGSSAGLMVAASAPYVAMMSVTAGQAELTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWIGSSAGLMVAASAPYVAMMSVTAGQAELTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPPVIAENRAELMILIAITNLLGQNTPAIAVNEAEYGENWAOQAAAMFYAAAT 300
Db 241 TAYGLTVPPPVIAENRAELMILIAITNLLGQNTPAIAVNEAEYGENWAOQAAAMFYAAAT 300
QY 301 ATATATLLPFEAPPEMTSAGGLLEQAAAEEASDTAAANQLMNNVPOALQQLAQTQGT 360
Db 301 ATATATLLPFEAPPEMTSAGGLLEQAAAEEASDTAAANQLMNNVPOALQQLAQTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAARQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSSGGLGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSLGSSGGLGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
QY 481 LTSLSAAERPGQMLGGLPVGQMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
Db 481 LTSLSAAERPGQMLGGLPVGQMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPDPDSAMVAQVGPQVNVINTKLGYNNAVAGAGTGIVIDPENGVLTNHVTIA 600
Db 541 QDRFADFPALPDPDSAMVAQVGPQVNVINTKLGYNNAVAGAGTGIVIDPENGVLTNHVTIA 600
QY 601 GATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Db 601 GATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
QY 661 SGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
Db 661 SGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
QY 721 VVGMTAAS 729
Db 721 VVGMTAAS 729

RESULT 8
US-10-369-983-14
; Sequence 14, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15

```
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 825
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB89F (MTB72F-MTI)
US-10-369-983-14

Query Match          99.8%; Score 3680; DB 15; Length 825;
Best Local Similarity 99.9%; Pred. No. 2.3e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGTAFILGLGVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGTAFILGLGVD 60
QY 61 NNGGARVQRVVGSAAPASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
Db 61 NNGGARVQRVVGSAAPASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
QY 121 TKSGGTRGNVTLAEGPPAEFMDVFGALPPEINSARMYAGPGSASLVAAQAQMDVSD 180
Db 121 TKSGGTRGNVTLAEGPPAEFMDVFGALPPEINSARMYAGPGSASLVAAQAQMDVSD 180
QY 181 FSAASAFQSVVWGLTVGSWIGSSAGLVAAASPVVAMWSVTAGQAEITAAQVRAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSWIGSSAGLVAAASPVVAMWSVTAGQAEITAAQVRAAAAYE 240
QY 241 TAYGLTVPPVPIAENRAELMLIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVPIAENRAELMLIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPPEEAPEMTSAGLLLEQAAAVEASDTAAANQLMNNVPAQLQALQPTQGT 360
Db 301 ATATATLLPPEEAPEMTSAGLLLEQAAAVEASDTAAANQLMNNVPAQLQALQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPQAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPQAAANQAVTPAARALP 480
QY 481 LTSLSAERGPQMLGGLPVQMGARAGGLSGVLVPRPVPYMPHSPAAGDIAPPALS 540
Db 481 LTSLSAERGPQMLGGLPVQMGARAGGLSGVLVPRPVPYMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVVDRDQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 600
Db 541 QDRFADFPALPLDPSAMVAQVGVVDRDQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 600
QY 601 GATDINAFSVSGQTYGVVDVVDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 660
Db 601 GATDINAFSVSGQTYGVVDVVDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 660
QY 661 SGQGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
Db 661 SGQGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
QY 721 VVGMNTAAS 729
Db 721 VVGMNTAAS 729
```

RESULT 9

US-10-369-983-13

; Sequence 13, Application US/10369983

; Publication No. US20030235593A1

; GENERAL INFORMATION:

```
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 875
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB89F (MTB72F-Erd14)
US-10-369-983-13
```

```
Query Match          99.8%; Score 3680; DB 15; Length 875;
Best Local Similarity 99.9%; Pred. No. 2.4e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGTAFILGLGVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGTAFILGLGVD 60
QY 61 NNGGARVQRVVGSAAPASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
Db 61 NNGGARVQRVVGSAAPASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
QY 121 TKSGGTRGNVTLAEGPPAEFMDVFGALPPEINSARMYAGPGSASLVAAQAQMDVSD 180
Db 121 TKSGGTRGNVTLAEGPPAEFMDVFGALPPEINSARMYAGPGSASLVAAQAQMDVSD 180
QY 181 FSAASAFQSVVWGLTVGSWIGSSAGLVAAASPVVAMWSVTAGQAEITAAQVRAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSWIGSSAGLVAAASPVVAMWSVTAGQAEITAAQVRAAAAYE 240
QY 241 TAYGLTVPPVPIAENRAELMLIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVPIAENRAELMLIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPPEEAPEMTSAGLLLEQAAAVEASDTAAANQLMNNVPAQLQALQPTQGT 360
Db 301 ATATATLLPPEEAPEMTSAGLLLEQAAAVEASDTAAANQLMNNVPAQLQALQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPQAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPQAAANQAVTPAARALP 480
QY 481 LTSLSAERGPQMLGGLPVQMGARAGGLSGVLVPRPVPYMPHSPAAGDIAPPALS 540
Db 481 LTSLSAERGPQMLGGLPVQMGARAGGLSGVLVPRPVPYMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVVDRDQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 600
Db 541 QDRFADFPALPLDPSAMVAQVGVVDRDQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 600
QY 601 GATDINAFSVSGQTYGVVDVVDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 660
Db 601 GATDINAFSVSGQTYGVVDVVDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 660
QY 661 SGQGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
Db 661 SGQGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
```


QY	541	QDRFADFPALP	LDPSAMVACVGPVVVNTKLGYNNAVAGAGTGV	LDPNGVVLNNH	600		
Db	541	QDRFADFPALP	LDPSAMVACVGPVVVNTKLGYNNAVAGAGTGV	LDPNGVVLNNH	600		
QY	601	GATDINAFSVG	SQTYGV	DVWGVDRTQD	VAVLQLRGAGGLPSAA	IGGVAVGPPVAMGN	660
Db	601	GATDINAFSVG	SQTYGV	DVWGVDRTQD	VAVLQLRGAGGLPSAA	IGGVAVGPPVAMGN	660
QY	661	SGGGGTTPRAV	PGRVVVALGQTV	QASDSLTCGABETL	NGLICFDAA	IOPGDSGGPVVNLGQ	720
Db	661	SGGGGTTPRAV	PGRVVVALGQTV	QASDSLTCGABETL	NGLICFDAA	IOPGDSGGPVVNLGQ	720
QY	721	VVGMTAA	S	729			
Db	721	VVGMTAA	S	729			
RESULT 11							
US-10-369-983-12							
; Sequence 12, Application US/10369983							
; Publication No. US20030235593A1							
; GENERAL INFORMATION:							
; APPLICANT: Skeiky, Yasir							
; APPLICANT: Guderian, Jeff							
; APPLICANT: Reed, Steven							
; APPLICANT: Corixa Corporation							
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis							
; FILE REFERENCE: 014058-009081US							
; CURRENT APPLICATION NUMBER: US/10/369,983							
; CURRENT FILING DATE: 2003-02-18							
; PRIOR APPLICATION NUMBER: US 60/357,351							
; PRIOR FILING DATE: 2002-02-15							
; NUMBER OF SEQ ID NOS: 22							
; SOFTWARE: PatentIn Ver. 2.1							
; SEQ ID NO 12							
; LENGTH: 930							
; TYPE: PRT							
; ORGANISM: Artificial Sequence							
; FEATURE:							
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein							
; OTHER INFORMATION: R95F (MTB72F-NAPS)							
US-10-369-983-12							
Query Match							
Best Local Similarity 99.8%; Score 3680; DB 15; Length 930;							
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;							
QY	1	MHHHHHTA	SDNFQLSGGGGFAIPICQAMA	TAGQIRSGGGSPTVHIGPTAF	LGLG	VVD	60
Db	1	MHHHHHTA	SDNFQLSGGGGFAIPICQAMA	TAGQIRSGGGSPTVHIGPTAF	LGLG	VVD	60
QY	61	NNGGARVQR	VWGSAPAAS	IGISTGDVITAVD	GPAPINSATAMADAL	NHHPGDV	120
Db	61	NNGGARVQR	VWGSAPAAS	IGISTGDVITAVD	GPAPINSATAMADAL	NHHPGDV	120
QY	121	TKSGGTGT	GNVTIAEGPP	PAEFMDVFGAL	PEIINSARMYAGP	GSASLVAAQ	180
Db	121	TKSGGTGT	GNVTIAEGPP	PAEFMDVFGAL	PEIINSARMYAGP	GSASLVAAQ	180
QY	181	FSAASAFQ	SVVWGLTVG	TSWIGSSAGL	MVAASP	YVAMWSVTAG	240
Db	181	FSAASAFQ	SVVWGLTVG	TSWIGSSAGL	MVAASP	YVAMWSVTAG	240
QY	241	TAYGLTV	PPVPIAENRAEL	MILLIATNLL	QNTPAIAVNEAEY	GEMWQDA	300
Db	241	TAYGLTV	PPVPIAENRAEL	MILLIATNLL	QNTPAIAVNEAEY	GEMWQDA	300
QY	301	ATATATILL	PFEEA	PEMTSAGGLI	PQAAAEAS	DTAAANQLMNNV	360
Db	301	ATATATILL	PFEEA	PEMTSAGGLI	PQAAAEAS	DTAAANQLMNNV	360
QY	361	PSSKLG	LWKTVSPHR	SPINNMWS	MANNNHMTN	TSNGSVMTN	420

Qy	721	VUGMNTAAS	729
Db	721	VUGMNTAAS	729
RESULT 10			
US-10-098-732A-65			
; Sequence 65, Application US/10098732A			
; Publication No. US20030175294A1			
; GENERAL INFORMATION:			
; APPLICANT: Skeiky, Yasir			
; APPLICANT: Brannon, Mark			
; APPLICANT: Guderian, Jeffrey			
; APPLICANT: Corixa Corporation			
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a			
; FILE REFERENCE: 014058-012010US			
; CURRENT APPLICATION NUMBER: US/10/098,732A			
; CURRENT FILING DATE: 2003-04-29			
; PRIOR APPLICATION NUMBER: US 60/275,837			
; PRIOR FILING DATE: 2001-03-13			
; NUMBER OF SEQ ID NOS: 80			
; SOFTWARE: Patentin Ver. 2.1			
; SEQ ID NO 65			
; LENGTH: 930			
; TYPE: PRT			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence:MTE72F-MAPS			
; OTHER INFORMATION: (r35f) fusion construct, TB MTE72F (Ral2-TbH9-Ra35)			
; OTHER INFORMATION: linked to Leishmania thiol-specific antioxidant			
; OTHER INFORMATION: (TSA or MAPS)			
US-10-098-732A-65			
Query Match 99.8%; Score 3680; DB 14; Length 930;			
Best Local Similarity 99.9%; Pred.No.2.6e-243; Indels 0; Gaps 0;			
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
Qy	1	MHHHHHTAASDNFQLSGGGGFAIPICQMAGIQAIRSGGGSPTVHI	GPTAFILGLGWD 60
Db	1	MEHHHHHTAASDNFQLSGGGGFAIPICQMAGIQAIRSGGGSPTVHI	GPTAFILGLGWD 60
Qy	61	NNGGARVQRVVGSAPAAASIGISTGDVITAVDGPINISATAMADALN	HGHHPGVISTWQ 120
Db	61	NNGGARVQRVVGSAPAAASIGISTGDVITAVDGPINISATAMADALN	HGHHPGVISTWQ 120
Qy	121	TKSGGTGTGNVTLAGPPAPFPWFDFGALPPEINSARMYAGPGSASLVA	AQQMWDSVASDL 180
Db	121	TKSGGTGTGNVTLAGPPAPFPWFDFGALPPEINSARMYAGPGSASLVA	AQQMWDSVASDL 180
Qy	181	FSAASAFAQSVVWGTLVGSSAGLMVAAAASPYVAMVSVTAGQAEITAA	QVRVAAAAAYE 240
Db	181	FSAASAFAQSVVWGTLVGSSAGLMVAAAASPYVAMVSVTAGQAEITAA	QVRVAAAAAYE 240
Qy	241	TAYGLTVPPPPIAENRAELMTLIATNLGONTPAIAVNEABYGEMWAQDA	AAAMEGYAAAT 300
Db	241	TAYGLTVPPPPIAENRAELMTLIATNLGONTPAIAVNEABYGEMWAQDA	AAAMEGYAAAT 300
Qy	301	ATAATATLLPFEEAPEMTSAGLLLEQAAAVEASDTAAANLMMNVPOALQ	LQAQTGTT 360
Db	301	ATAATATLLPFEEAPEMTSAGLLLEQAAAVEASDTAAANLMMNVPOALQ	LQAQTGTT 360
Qy	361	PSSKLGIGLWKTVSPHRSPISNNWSMANNHMTNSGVSMNTLTLSMLKGPA	APAARQAVQ 420
Db	361	PSSKLGIGLWKTVSPHRSPISNNWSMANNHMTNSGVSMNTLTLSMLKGPA	APAARQAVQ 420
Qy	421	TAAQNGVRAMSLSGSSLGGSLGGCVAAANTGRAASVGSLSVPQAWAANAQ	AVTPAAALP 480
Db	421	TAAQNGVRAMSLSGSSLGGSLGGCVAAANTGRAASVGSLSVPQAWAANAQ	AVTPAAALP 480
Qy	481	LTSLTSAAERPGOMLGGLPVQMGARAGGGLSGVLRVPPRPYVNPISPAAGDI	APPALS 540
Db	481	LTSLTSAAERPGOMLGGLPVQMGARAGGGLSGVLRVPPRPYVNPISPAAGDI	APPALS 540

361 PSSKLGGLWKTVSPHRSPISNWVSMANNHMSMTNSGVSMNTNTLSSMLKGFAPAAAQAVQ 420
421 TAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVQAWAANAQAVTPAARALP 480
421 TAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVQAWAANAQAVTPAARALP 480
481 LTSLSAERPGQMLGGLPYQMGARAGGGLSGVLRVPPRYMHPSPAAAGDIAPPALS 540
481 LTSLSAERPGQMLGGLPYQMGARAGGGLSGVLRVPPRYMHPSPAAAGDIAPPALS 540
541 QDRFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVGAGTGIVDPNGVLTNNHVIA 600
541 QDRFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVGAGTGIVDPNGVLTNNHVIA 600
601 GATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGN 660
601 GATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGN 660
661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDSGGPVVWGLGQ 720
661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDSGGPVVWGLGQ 720
721 VVGMMNTAAS 729
721 VVGMMNTAAS 729

RESULT 12

US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85b)
US-10-369-983-18

Query Match 99.8%; Score 3680; DB 15; Length 1016;
Best Local Similarity 99.9%; Pred. No. 2.9e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPQAWAIAQIRSGGSPVHIGTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQLSQGGGFAIPQAWAIAQIRSGGSPVHIGTAFGLGVVD 60
QY 61 NNGNGARVQRVVGSAAPASLIGISTGDIITAVDGAPINSATAMADALNGHHPGDVISTWQ 120
DB 61 NNGNGARVQRVVGSAAPASLIGISTGDIITAVDGAPINSATAMADALNGHHPGDVISTWQ 120
QY 121 TKSGGTTRTGNVTLAEGPPAEFMVDFGALPPEINSAARMYAGPGSASLVAAQWDSVASDL 180
DB 121 TKSGGTTRTGNVTLAEGPPAEFMVDFGALPPEINSAARMYAGPGSASLVAAQWDSVASDL 180
QY 181 FSAASAFQSVWGLTVGSMIGSSAGLWVAASPYVWMSVTAGQABLTAQVRVAAAAYE 240
DB 181 FSAASAFQSVWGLTVGSMIGSSAGLWVAASPYVWMSVTAGQABLTAQVRVAAAAYE 240

QY 241 TAYGLTVPPPVAENRAELMLIATNLLGONTPAIVANAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPPVAENRAELMLIATNLLGONTPAIVANAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLPPFEAPEMTSAGGLEQAAAABEASDTAAANQLMNNVPQALQOLAQPTQGT 360
DB 301 ATATATLPPFEAPEMTSAGGLEQAAAABEASDTAAANQLMNNVPQALQOLAQPTQGT 360
QY 361 PSSKLGGLWKTVSPHRSPISNWVSMANNHMSMTNSGVSMNTNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTVSPHRSPISNWVSMANNHMSMTNSGVSMNTNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVQAWAANAQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVQAWAANAQAVTPAARALP 480
QY 481 LTSLSAERPGQMLGGLPYQMGARAGGGLSGVLRVPPRYMHPSPAAAGDIAPPALS 540
DB 481 LTSLSAERPGQMLGGLPYQMGARAGGGLSGVLRVPPRYMHPSPAAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVGAGTGIVDPNGVLTNNHVIA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVGAGTGIVDPNGVLTNNHVIA 600
QY 601 GATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGN 660
DB 601 GATDINAFSVSGGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDSGGPVVWGLGQ 720
DB 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDSGGPVVWGLGQ 720
QY 721 VVGMMNTAAS 729
DB 721 VVGMMNTAAS 729

RESULT 13

US-10-369-983-17
; Sequence 17, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1022
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB102tm2F (MTB102FTW2, MTB72F-HTCC#1)
US-10-369-983-17

Query Match 99.8%; Score 3680; DB 15; Length 1022;
Best Local Similarity 99.9%; Pred. No. 3e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPQAWAIAQIRSGGSPVHIGTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQLSQGGGFAIPQAWAIAQIRSGGSPVHIGTAFGLGVVD 60
QY 61 NNGNGARVQRVVGSAAPASLIGISTGDIITAVDGAPINSATAMADALNGHHPGDVISTWQ 120

Db 61 NNGNGARVQVVGSAFAAASLGISTGDIITAVDGPINSATAMADALNGHHPGDVLSVTWQ 120
QY 121 TKSGGTRTGNVTLAEQPPAEFVDFGALPPEINSARMYAGPGSASIVAAQWDSVASDL 180
Db 121 TKSGGTRTGNVTLAEQPPAEFVDFGALPPEINSARMYAGPGSASIVAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWISAGLMVAAAPYVAMSVTAGQAELETAQVRAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWISAGLMVAAAPYVAMSVTAGQAELETAQVRAAAAYE 240
QY 241 TAYGLTVPPVIAENRAELMILIAINLLGONTPTAIVNEAEYGENWAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVIAENRAELMILIAINLLGONTPTAIVNEAEYGENWAQDAAMFGYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGGLEQAAAEEASDTAAANQLMNNVPOALQQLAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGGLEQAAAEEASDTAAANQLMNNVPOALQQLAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNNVSMANNHMSMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNNVSMANNHMSMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSSLSGSSGLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLSGSSGLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALP 480
QY 481 LTSLSAAERPGQMLGGLPVQMGARAGGSLGVLRVPPRYMPPHSPAAGDIAPPALS 540
Db 481 LTSLSAAERPGQMLGGLPVQMGARAGGSLGVLRVPPRYMPPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVVNTIKLYNNAVGAGTGIVIDPNGVLTNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGVVNTIKLYNNAVGAGTGIVIDPNGVLTNNHVA 600
QY 601 GATDINAFSVSGSQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVMGN 660
Db 601 GATDINAFSVSGSQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVMGN 660
QY 661 SGGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
Db 661 SGGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
QY 721 VVGWMTAAS 729
Db 721 VVGWMTAAS 729

RESULT 14
US-10-369-983-16
; Sequence 16, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guiderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB114F (MTB72P-mtCC#2)
US-10-369-983-16

Query Match 99.8%; Score 3680; DB 15; Length 1154;
Best Local Similarity 99.9%; Pred. No. 3.5e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFQLSQGGQFAIPQOAMAIAGQIRSGGSPVTHIGTAFGLGVD 60
Db 1 MHHHHHTAASDNFQLSQGGQFAIPQOAMAIAGQIRSGGSPVTHIGTAFGLGVD 60
QY 61 NNGNGARVQVVGSAFAAASLGISTGDIITAVDGPINSATAMADALNGHHPGDVLSVTWQ 120
Db 61 NNGNGARVQVVGSAFAAASLGISTGDIITAVDGPINSATAMADALNGHHPGDVLSVTWQ 120
QY 121 TKSGGTRTGNVTLAEQPPAEFVDFGALPPEINSARMYAGPGSASIVAAQWDSVASDL 180
Db 121 TKSGGTRTGNVTLAEQPPAEFVDFGALPPEINSARMYAGPGSASIVAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWISAGLMVAAAPYVAMSVTAGQAELETAQVRAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWISAGLMVAAAPYVAMSVTAGQAELETAQVRAAAAYE 240
QY 241 TAYGLTVPPVIAENRAELMILIAINLLGONTPTAIVNEAEYGENWAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVIAENRAELMILIAINLLGONTPTAIVNEAEYGENWAQDAAMFGYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGGLEQAAAEEASDTAAANQLMNNVPOALQQLAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGGLEQAAAEEASDTAAANQLMNNVPOALQQLAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNNVSMANNHMSMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNNVSMANNHMSMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSSLSGSSGLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLSGSSGLGGGVAANLGRAASVGSLSVPQAWAAANQAVTPAARALP 480
QY 481 LTSLSAAERPGQMLGGLPVQMGARAGGSLGVLRVPPRYMPPHSPAAGDIAPPALS 540
Db 481 LTSLSAAERPGQMLGGLPVQMGARAGGSLGVLRVPPRYMPPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVVNTIKLYNNAVGAGTGIVIDPNGVLTNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGVVNTIKLYNNAVGAGTGIVIDPNGVLTNNHVA 600
QY 601 GATDINAFSVSGSQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVMGN 660
Db 601 GATDINAFSVSGSQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVMGN 660
QY 661 SGGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
Db 661 SGGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
QY 721 VVGWMTAAS 729
Db 721 VVGWMTAAS 729

RESULT 15
US-09-886-349A-18
; Sequence 18, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20

;; PRIOR APPLICATION NUMBER: US 60/265,737
;; PRIOR FILING DATE: 2001-02-01
;; NUMBER OF SEQ ID NOS: 50
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 18
;; LENGTH: 729
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence.MTB72PMutSA
;; OTHER INFORMATION: (Ra12-TbHp-Ra35MutSA)
US-09-886-349A-18

Query Match		99.8%	Score 3677	DB 12	Length 729
Best Local Similarity		99.7%	Pred. No. 3.1e-243		
Matches 727	Conservative 1	Mismatches 1	Indels 0	Gaps 0	
QY	1	MHHHHHTAASDNFQSGGGGFAIPIGQAWAIAAGQIRSGGSPTHIGPTAFGLGVVD	60		
Db	1	MHHHHHTAASDNFQSGGGGFAIPIGQAWAIAAGQIRSGGSPTHIGPTAFGLGVVD	60		
QY	61	NNNGARVQVRVVGSAAPASLGISTGDIVITAVDGAPINSATAMADALNGHHPGDIVISVTWQ	120		
Db	61	NNNGARVQVRVVGSAAPASLGISTGDIVITAVDGAPINSATAMADALNGHHPGDIVISVTWQ	120		
QY	121	TKSGGTRTGNVTLAEGPPAEPFWDGALPPEINSARMYAGPGSASLVAAAQMWDSVASDL	180		
Db	121	TKSGGTRTGNVTLAEGPPAEPFWDGALPPEINSARMYAGPGSASLVAAAQMWDSVASDL	180		
QY	181	FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMWSVTAGQAEITAAQVRAAAAYE	240		
Db	181	FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMWSVTAGQAEITAAQVRAAAAYE	240		
QY	241	TAYGLTVPPVPVIAENRAELMILITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT	300		
Db	241	TAYGLTVPPVPVIAENRAELMILITNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT	300		
QY	301	ATATATLLPREEAEMTSAGLLEQAAVEASDTAAANOLMNNVPOALQOLAQPTQGT	360		
Db	301	ATATATLLPREEAEMTSAGLLEQAAVEASDTAAANOLMNNVPOALQOLAQPTQGT	360		
QY	361	PSSKLGGLWKTVPSPHRSPI SNMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ	420		
Db	361	PSSKLGGLWKTVPSPHRSPI SNMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ	420		
QY	421	TAAQNGVRAMSSLGSSLGSGGVAANI GRAASVGSLSVPOAWAANQAVTPAARALP	480		
Db	421	TAAQNGVRAMSSLGSSLGSGGVAANI GRAASVGSLSVPOAWAANQAVTPAARALP	480		
QY	481	LTSITSAAERGPQMLGGLPVQMGARAGGLSGVLVRPPRPVYMPHSPAAGDIAPPALS	540		
Db	481	LTSITSAAERGPQMLGGLPVQMGARAGGLSGVLVRPPRPVYMPHSPAAGDIAPPALS	540		
QY	541	QDRFADFPALPLDPSAMVAQVGQVNNINTKLYNNNAVAGTGVITDPNGVLTNNHVIA	600		
Db	541	QDRFADFPALPLDPSAMVAQVGQVNNINTKLYNNNAVAGTGVITDPNGVLTNNHVIA	600		
QY	601	GATDINAFSVGSGTGVVDVVGVDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMGN	660		
Db	601	GATDINAFSVGSGTGVVDVVGVDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMGN	660		
QY	661	SGCGGTTPRAVGRVVALGQTVQASDSLTAETLNLGIQFDAAIQPGSGGPFVNGLQ	720		
Db	661	SGCGGTTPRAVGRVVALGQTVQASDSLTAETLNLGIQFDAAIQPGSGGPFVNGLQ	720		
QY	721	VVGWNTAAS 729			
Db	721	VVGWNTAAS 729			

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 67.5195 Seconds
(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-18
Perfect score: 3685
Sequence: 1 MHHHHHTAASDNFOLSGGG.....AGGPVYVGLGVGMNTAAS 729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

ALIGNMENTS

RESULT 1

US-09-886-349A-18
; Sequence 18, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886.349A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; PRIOR FILING DATE: 2000-06-20
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72FWtUSA
; OTHER INFORMATION: (Ra12-TbHp-Ra3SWtUSA)
US-09-886-349A-18

Query Match 100.0%; Score 3685; DB 12; Length 729;
Best Local Similarity 100.0%; Pred. No. 3.4e-239;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFOLSGGGGFAIPICQAVAIAGIIRSGGSPVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFOLSGGGGFAIPICQAVAIAGIIRSGGSPVHIGPTAFGLGVVD 60
QY 61 NNGNGARVQRVGVGSPAASLGIISTGDVITAVDGFINSATAMADALNGHHPGDVTSVTWQ 120

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3685	100.0	729	12	US-09-886-349A-18
2	3685	100.0	729	14	US-10-098-732A-18
3	3685	100.0	729	15	US-10-369-983-22
4	3682	99.9	729	15	US-10-369-983-21
5	3682	99.9	813	15	US-10-369-983-15
6	3682	99.9	825	15	US-10-369-983-14
7	3682	99.9	875	15	US-10-369-983-13
8	3682	99.9	930	14	US-10-098-732A-65
9	3682	99.9	930	15	US-10-369-983-12
10	3682	99.9	1016	15	US-10-369-983-18
11	3682	99.9	1022	15	US-10-369-983-17
12	3682	99.9	1154	15	US-10-369-983-16
13	3677	99.8	729	9	US-09-287-849-2
14	3677	99.8	729	12	US-09-886-349A-18
15	3677	99.8	729	14	US-10-359-450-2

Db 61 NNGNGARVQRVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNHHHPGDIIVSTWQ 120
Qy 121 TKSGGTRTGNVTLAEGPPAEPMVDFGALPPEINARMYAGPGSASLVAAQMWDSVASDL 180
Db 121 TKSGGTRTGNVTLAEGPPAEPMVDFGALPPEINARMYAGPGSASLVAAQMWDSVASDL 180
Qy 181 FSAASAFQSVVWGLTVGSSAGLVAAASPVYAMSVTAGOAEITAAQVRAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSSAGLVAAASPVYAMSVTAGOAEITAAQVRAAAAYE 240
Qy 241 TAYGLTVPPPIAENRAELMILIAINLLGQNTPAIAVNEAEYEGMWAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPPIAENRAELMILIAINLLGQNTPAIAVNEAEYEGMWAQDAAMFGYAAAT 300
Qy 301 ATATATLLPFEAEPMTSAGGLLEQAAAVBEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Db 301 ATATATLLPFEAEPMTSAGGLLEQAAAVBEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Qy 361 PSSKLGGLWKTVPSPHRSPISNMVSMANNHSMNTNSGVSWTNTLSSMLKGFAPAAAQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNMVSMANNHSMNTNSGVSWTNTLSSMLKGFAPAAAQAVQ 420
Qy 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPQAAWAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPQAAWAAANQAVTPAARALP 480
Qy 481 LTSLSAERPGQMLGGLPVQNGARAGGLSGVLVPRPYPYMPHSPPAGDIAPPALS 540
Db 481 LTSLSAERPGQMLGGLPVQNGARAGGLSGVLVPRPYPYMPHSPPAGDIAPPALS 540
Qy 541 QDRFADFPALPLDPSAMVAQVGVQVNVINTKLYNNVAVGAGTGIVDPNGVLTNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGVQVNVINTKLYNNVAVGAGTGIVDPNGVLTNNHVA 600
Qy 601 GATDINAFSVGSGGTGYVDVVGVDRTODVAVLQIRGAGGLPSAAIGGVAVGEPVAVMGN 660
Db 601 GATDINAFSVGSGGTGYVDVVGVDRTODVAVLQIRGAGGLPSAAIGGVAVGEPVAVMGN 660
Qy 661 SGGGGTTPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQPDAGGPPVNVGLGQ 720
Db 661 SGGGGTTPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQPDAGGPPVNVGLGQ 720
Qy 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

RESULT 2

US-10-098-732A-18
; Sequence 18, Application US/10098732A
; Publication No. US2003017529A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098, 732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275, 837
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB72FmutSa
; OTHER INFORMATION: (Ra12-TbH9-Ra15mutSa)

Query Match 100.0%; Score 3685; DB 14; Length 729;
Best Local Similarity 100.0%; Pred. No. 3.46-239;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHTAASDNFOLSQGGGFAIPICQAMAIAGQIRSGGSPYVHIGTAFILGLVVD 60
Db 1 MHHHHHTAASDNFOLSQGGGFAIPICQAMAIAGQIRSGGSPYVHIGTAFILGLVVD 60
Qy 61 NNGNGARVQRVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNHHHPGDIIVSTWQ 120
Db 61 NNGNGARVQRVVGSAAPASLGISTGDTVITAVDGAIPINSATAMADALNHHHPGDIIVSTWQ 120
Qy 121 TKSGGTRTGNVTLAEGPPAEPMVDFGALPPEINARMYAGPGSASLVAAQMWDSVASDL 180
Db 121 TKSGGTRTGNVTLAEGPPAEPMVDFGALPPEINARMYAGPGSASLVAAQMWDSVASDL 180
Qy 181 FSAASAFQSVVWGLTVGSSAGLVAAASPVYAMSVTAGOAEITAAQVRAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSSAGLVAAASPVYAMSVTAGOAEITAAQVRAAAAYE 240
Qy 241 TAYGLTVPPPIAENRAELMILIAINLLGQNTPAIAVNEAEYEGMWAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPPIAENRAELMILIAINLLGQNTPAIAVNEAEYEGMWAQDAAMFGYAAAT 300
Qy 301 ATATATLLPFEAEPMTSAGGLLEQAAAVBEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Db 301 ATATATLLPFEAEPMTSAGGLLEQAAAVBEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Qy 361 PSSKLGGLWKTVPSPHRSPISNMVSMANNHSMNTNSGVSWTNTLSSMLKGFAPAAAQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNMVSMANNHSMNTNSGVSWTNTLSSMLKGFAPAAAQAVQ 420
Qy 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPQAAWAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPQAAWAAANQAVTPAARALP 480
Qy 481 LTSLSAERPGQMLGGLPVQNGARAGGLSGVLVPRPYPYMPHSPPAGDIAPPALS 540
Db 481 LTSLSAERPGQMLGGLPVQNGARAGGLSGVLVPRPYPYMPHSPPAGDIAPPALS 540
Qy 541 QDRFADFPALPLDPSAMVAQVGVQVNVINTKLYNNVAVGAGTGIVDPNGVLTNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGVQVNVINTKLYNNVAVGAGTGIVDPNGVLTNNHVA 600
Qy 601 GATDINAFSVGSGGTGYVDVVGVDRTODVAVLQIRGAGGLPSAAIGGVAVGEPVAVMGN 660
Db 601 GATDINAFSVGSGGTGYVDVVGVDRTODVAVLQIRGAGGLPSAAIGGVAVGEPVAVMGN 660
Qy 661 SGGGGTTPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQPDAGGPPVNVGLGQ 720
Db 661 SGGGGTTPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQPDAGGPPVNVGLGQ 720
Qy 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

RESULT 3

US-10-369-983-22
; Sequence 22, Application US/10369983
; Publication No. US2003023559A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369, 983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357, 351

```
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 22
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: mutated
; OTHER INFORMATION: MTB72FmutSA (Mtb72f-mur-SA)
US-10-369-983-22

Query Match      100.0%; Score 3685; DB 15; Length 729;
Best Local Similarity 100.0%; Pred. No. 3.4e-239;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQSQGGGQFAIPGQAMAIAGQIRSGGGSPVHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQSQGGGQFAIPGQAMAIAGQIRSGGGSPVHIGPTAFGLGVVD 60
QY 61 NNGNGARQVRVGSAPAAASLGISTGCVITAVDGPAPINSATAMADALNGHHPGDVSVTWQ 120
DB 61 NNGNGARQVRVGSAPAAASLGISTGCVITAVDGPAPINSATAMADALNGHHPGDVSVTWQ 120
QY 121 TKSOGTRTGNVTLAEGPPAEFVDFGALPPEINARMYAGPGSASLVAAAQWDSVASDL 180
DB 121 TKSOGTRTGNVTLAEGPPAEFVDFGALPPEINARMYAGPGSASLVAAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWISAGLMVAAAAPYVAMSVTAGQAEELTAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSGWISAGLMVAAAAPYVAMSVTAGQAEELTAQVRVAAAAYE 240
QY 241 TAYGLTVPPPIAENRAELMILITNLLQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPPIAENRAELMILITNLLQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPPEAPEMTSAGLLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGT 360
DB 301 ATATATLLPPEAPEMTSAGLLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGT 360
QY 361 PSSKLGGLWKTIVSPHRSPISNNVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTIVSPHRSPISNNVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALP 480
QY 481 LTSLSAAERGPQMLGGLPVQMGARAGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
DB 481 LTSLSAAERGPQMLGGLPVQMGARAGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGTGIIVDPNGVLTNNHVIA 600
DB 541 QDRFADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGTGIIVDPNGVLTNNHVIA 600
QY 601 GATDINAFSVSGQTYGVVDVDTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
DB 601 GATDINAFSVSGQTYGVVDVDTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
QY 661 SGGGGTTPRVPGRVVALGQTVQASDSLTAAGETLNGLIQFDAAIQPDAGGPPVNGLGQ 720
DB 661 SGGGGTTPRVPGRVVALGQTVQASDSLTAAGETLNGLIQFDAAIQPDAGGPPVNGLGQ 720
QY 721 VVGNTAAS 729
DB 721 VVGNTAAS 729

; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 21
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB72F
US-10-369-983-21

Query Match      99.9%; Score 3682; DB 15; Length 729;
Best Local Similarity 99.9%; Pred. No. 5.4e-239;
Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQSQGGGQFAIPGQAMAIAGQIRSGGGSPVHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQSQGGGQFAIPGQAMAIAGQIRSGGGSPVHIGPTAFGLGVVD 60
QY 61 NNGNGARQVRVGSAPAAASLGISTGCVITAVDGPAPINSATAMADALNGHHPGDVSVTWQ 120
DB 61 NNGNGARQVRVGSAPAAASLGISTGCVITAVDGPAPINSATAMADALNGHHPGDVSVTWQ 120
QY 121 TKSOGTRTGNVTLAEGPPAEFVDFGALPPEINARMYAGPGSASLVAAAQWDSVASDL 180
DB 121 TKSOGTRTGNVTLAEGPPAEFVDFGALPPEINARMYAGPGSASLVAAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWISAGLMVAAAAPYVAMSVTAGQAEELTAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSGWISAGLMVAAAAPYVAMSVTAGQAEELTAQVRVAAAAYE 240
QY 241 TAYGLTVPPPIAENRAELMILITNLLQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPPIAENRAELMILITNLLQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPPEAPEMTSAGLLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGT 360
DB 301 ATATATLLPPEAPEMTSAGLLEQAAAVEEASDTAAANQLMNNVPOALQQLAQPTQGT 360
QY 361 PSSKLGGLWKTIVSPHRSPISNNVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTIVSPHRSPISNNVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAANAQAVTPAARALP 480
QY 481 LTSLSAAERGPQMLGGLPVQMGARAGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
DB 481 LTSLSAAERGPQMLGGLPVQMGARAGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGTGIIVDPNGVLTNNHVIA 600
DB 541 QDRFADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGTGIIVDPNGVLTNNHVIA 600
QY 601 GATDINAFSVSGQTYGVVDVDTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
DB 601 GATDINAFSVSGQTYGVVDVDTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
QY 661 SGGGGTTPRVPGRVVALGQTVQASDSLTAAGETLNGLIQFDAAIQPDAGGPPVNGLGQ 720
DB 661 SGGGGTTPRVPGRVVALGQTVQASDSLTAAGETLNGLIQFDAAIQPDAGGPPVNGLGQ 720
QY 721 VVGNTAAS 729
DB 721 VVGNTAAS 729

RESULT 4
US-10-369-983-21
; Sequence 21, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
```

Db 721 VVGMNTAAS 729
 |||||
 RESULT 5
 US-10-369-983-15
 ; Sequence 15, Application US/10369983
 ; Publication No. US20030235593A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Guderian, Jeff
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009081US
 ; CURRENT APPLICATION NUMBER: US/10/369,983
 ; CURRENT FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: US 60/357,351
 ; PRIOR FILING DATE: 2002-02-15
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 15
 ; LENGTH: 813
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
 ; OTHER INFORMATION: MTB81F (MTB72F-DPV)
 US-10-369-983-15

Query Match 99.9%; Score 3682; DB 15; Length 813;
 Best Local Similarity 99.9%; Pred. No. 6.2e-239;
 Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHHTAASDNFQLSQGGGFAIPICQAMAIAGQIRSGGSPTHVIGTAFILGLGVD 60
 Db 1 MHHHHHTAASDNFQLSQGGGFAIPICQAMAIAGQIRSGGSPTHVIGTAFILGLGVD 60
 QY 61 NNGNGARVQVRVGSAPAAASLGISTGVDITAVDGAPINSATAMADALNGHHPGDVISTWQ 120
 Db 61 NNGNGARVQVRVGSAPAAASLGISTGVDITAVDGAPINSATAMADALNGHHPGDVISTWQ 120
 QY 121 TKSGLTGNVTTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDL 180
 Db 121 TKSGLTGNVTTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDL 180
 QY 181 FSAASAFQSVVWGLTVGSGAGLMVAASPVVAWMSVTAGQAELETAQVRVAAAAYE 240
 Db 181 FSAASAFQSVVWGLTVGSGAGLMVAASPVVAWMSVTAGQAELETAQVRVAAAAYE 240
 QY 241 TAYGLTVPPPVIAENRAELMILITATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 Db 241 TAYGLTVPPPVIAENRAELMILITATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 QY 301 ATATATLLPPEEAPMTSAGLLLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGT 360
 Db 301 ATATATLLPPEEAPMTSAGLLLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGT 360
 QY 361 PSSKLGGLWKTVPSPHSPISNMVSMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
 Db 361 PSSKLGGLWKTVPSPHSPISNMVSMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
 QY 421 TAAQNGVRAMSSLGSSLGSGGCGGVAANLGRAASVGSISVFOQAAANQAVTPAARALP 480
 Db 421 TAAQNGVRAMSSLGSSLGSGGCGGVAANLGRAASVGSISVFOQAAANQAVTPAARALP 480
 QY 481 LTSLTSAAREGPQMLGGI:PVGMQAGARAGGILSGVLRVPPVPMPSHSPAAGDIAPPAL 540
 Db 481 LTSLTSAAREGPQMLGGI:PVGMQAGARAGGILSGVLRVPPVPMPSHSPAAGDIAPPAL 540
 QY 541 QDRFADFPALPDPDSAMVAQVGVVNVINTKLGYNNAVAGAGTGIVDPNGVVLTNHVA 600
 Db 541 QDRFADFPALPDPDSAMVAQVGVVNVINTKLGYNNAVAGAGTGIVDPNGVVLTNHVA 600

QY 601 GATDINAFSVGSGQTYGVDVVGVDRTQDVAVQLRAGAGLPSAAIGGGVAVGCEPVVAMGN 660
 Db 601 GATDINAFSVGSGQTYGVDVVGVDRTQDVAVQLRAGAGLPSAAIGGGVAVGCEPVVAMGN 660
 QY 661 SGGGGTTPRAVGRVVAIGQTVQASDSLTGABETLNGLIQPDAAIOPDAGAGPVVNGLQ 720
 Db 661 SGGGGTTPRAVGRVVAIGQTVQASDSLTGABETLNGLIQPDAAIOPDAGAGPVVNGLQ 720
 QY 721 VVGMNTAAS 729
 |||||
 Db 721 VVGMNTAAS 729
 |||||
 RESULT 6
 US-10-369-983-14
 ; Sequence 14, Application US/10369983
 ; Publication No. US20030235593A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Guderian, Jeff
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009081US
 ; CURRENT APPLICATION NUMBER: US/10/369,983
 ; CURRENT FILING DATE: 2003-02-18
 ; PRIOR APPLICATION NUMBER: US 60/357,351
 ; PRIOR FILING DATE: 2002-02-15
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 14
 ; LENGTH: 825
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
 ; OTHER INFORMATION: MTB83F (MTB72F-MTI)
 US-10-369-983-14

Query Match 99.9%; Score 3682; DB 15; Length 825;
 Best Local Similarity 99.9%; Pred. No. 6.3e-239;
 Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHHHHHTAASDNFQLSQGGGFAIPICQAMAIAGQIRSGGSPTHVIGTAFILGLGVD 60
 Db 1 MHHHHHTAASDNFQLSQGGGFAIPICQAMAIAGQIRSGGSPTHVIGTAFILGLGVD 60
 QY 61 NNGNGARVQVRVGSAPAAASLGISTGVDITAVDGAPINSATAMADALNGHHPGDVISTWQ 120
 Db 61 NNGNGARVQVRVGSAPAAASLGISTGVDITAVDGAPINSATAMADALNGHHPGDVISTWQ 120
 QY 121 TKSGLTGNVTTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDL 180
 Db 121 TKSGLTGNVTTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDL 180
 QY 181 FSAASAFQSVVWGLTVGSGAGLMVAASPVVAWMSVTAGQAELETAQVRVAAAAYE 240
 Db 181 FSAASAFQSVVWGLTVGSGAGLMVAASPVVAWMSVTAGQAELETAQVRVAAAAYE 240
 QY 241 TAYGLTVPPPVIAENRAELMILITATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 Db 241 TAYGLTVPPPVIAENRAELMILITATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 QY 301 ATATATLLPPEEAPMTSAGLLLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGT 360
 Db 301 ATATATLLPPEEAPMTSAGLLLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGT 360
 QY 361 PSSKLGGLWKTVPSPHSPISNMVSMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
 Db 361 PSSKLGGLWKTVPSPHSPISNMVSMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
 QY 421 TAAQNGVRAMSSLGSSLGSGGCGGVAANLGRAASVGSISVFOQAAANQAVTPAARALP 480


```

Db      421  TAAQNGVRAMSLGSLGSLGGVAAANLGRASVGLSVPOQAAANQAVTPAARALP  480
Qy      481  LLSLSAAERGGQMLGGLPVQMGARAGGGLSVGLRVPRPYVMPHSPAAAGDIAPPALS  540
Db      481  LLSLSAAERGGQMLGGLPVQMGARAGGGLSVGLRVPRPYVMPHSPAAAGDIAPPALS  540-
Qy      541  QRFADFPALPLDPSAMVAQVGPQVNNINIKLGYNNAVGAGTGVIVDPNGVVLTNHHVIA  600
Db      541  QRFADFPALPLDPSAMVAQVGPQVNNINIKLGYNNAVGAGTGVIVDPNGVVLTNHHVIA  600
Qy      601  GATDINAFVSGSGQTYGVVDVGYDRTQDVAVLQLRGAGGLPSAAIIGGGVAVGEPVVMGN  660
Db      601  GATDINAFVSGSGQTYGVVDVGYDRTQDVAVLQLRGAGGLPSAAIIGGGVAVGEPVVMGN  660
Qy      661  SGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQ  720
Db      661  SGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQ  720
Qy      721  VVGMMNTAAS 729
Db      721  VVGMMNTAAS 729

```

```

RESULT 7
US-10-369-983-13
; Sequence 13, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 875
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB89F (MTB72F-Erd14)
US-10-369-983-13

```

```

Query Match      99.9%; Score 3682; DB 15; Length 875;
Best Local Similarity 99.9%; Pred. No. 6.8e-239;
Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1  MHHHHHTAASDNFQSLGSGGQFAIPQGMATAGQIRSGGSPVTHIGPTAFGLGVVD  60
Db      1  MHHHHHTAASDNFQSLGSGGQFAIPQGMATAGQIRSGGSPVTHIGPTAFGLGVVD  60
Qy      61  NNGNGARVQRVGSAPASIGISTGDIVITAVDGAPINSATAMADALNGHPHGPDISVTWQ  120
Db      61  NNGNGARVQRVGSAPASIGISTGDIVITAVDGAPINSATAMADALNGHPHGPDISVTWQ  120
Qy      121  TKSGGTRTGNVTLAEGPPAEFMDVFGALPPEINSARYAGPGSASLVAAQWMDSVASDL  180
Db      121  TKSGGTRTGNVTLAEGPPAEFMDVFGALPPEINSARYAGPGSASLVAAQWMDSVASDL  180
Qy      181  FSAASAFQSVVWGLTVGSGWIGSSAGILMVAAASPYVAMSVTAGCAELTAQVEVRAAAVE  240
Db      181  FSAASAFQSVVWGLTVGSGWIGSSAGILMVAAASPYVAMSVTAGCAELTAQVEVRAAAVE  240
Qy      241  TAYGLTVPPVIAENRAELMILIAITNLGQNTPAIAVNEAEYGEYMAQDAAMFGYAAAT  300
Db      241  TAYGLTVPPVIAENRAELMILIAITNLGQNTPAIAVNEAEYGEYMAQDAAMFGYAAAT  300

```

```

Qy      301  ATATATLLPFEAPEMTSAGLLLEQAAAVEASDTAAANOLMNNVPOALQOLAQPTGTT  360
Db      301  ATATATLLPFEAPEMTSAGLLLEQAAAVEASDTAAANOLMNNVPOALQOLAQPTGTT  360
Qy      361  PSSKLGGLWKTIVSPHRSPISNNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAAQAVQ  420
Db      361  PSSKLGGLWKTIVSPHRSPISNNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAAQAVQ  420
Qy      421  TAAQNGVRAMSLGSLGSGGLGGVAAANLGRASVGLSVPOQAAANQAVTPAARALP  480
Db      421  TAAQNGVRAMSLGSLGSGGLGGVAAANLGRASVGLSVPOQAAANQAVTPAARALP  480
Qy      481  LTSLSAAERGGQMLGGLPVQMGARAGGGLSVGLRVPRPYVMPHSPAAAGDIAPPALS  540
Db      481  LTSLSAAERGGQMLGGLPVQMGARAGGGLSVGLRVPRPYVMPHSPAAAGDIAPPALS  540
Qy      541  QRFADFPALPLDPSAMVAQVGPQVNNINIKLGYNNAVGAGTGVIVDPNGVVLTNHHVIA  600
Db      541  QRFADFPALPLDPSAMVAQVGPQVNNINIKLGYNNAVGAGTGVIVDPNGVVLTNHHVIA  600
Qy      601  GATDINAFVSGSGQTYGVVDVGYDRTQDVAVLQLRGAGGLPSAAIIGGGVAVGEPVVMGN  660
Db      601  GATDINAFVSGSGQTYGVVDVGYDRTQDVAVLQLRGAGGLPSAAIIGGGVAVGEPVVMGN  660
Qy      661  SGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPCDAGGPVNGLGQ  720
Db      661  SGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPCDAGGPVNGLGQ  720
Qy      721  VVGMMNTAAS 729
Db      721  VVGMMNTAAS 729

RESULT 8
US-10-098-732A-65
; Sequence 65, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB72F-MAPS
; OTHER INFORMATION: (r95f) fusion construct, TB MTB72F (Ra12-TbH9-Ra35)
; OTHER INFORMATION: linked to Leishmania thiol-specific antioxidant
; OTHER INFORMATION: (TSA or MAPS)
US-10-098-732A-65

Query Match      99.9%; Score 3682; DB 14; Length 930;
Best Local Similarity 99.9%; Pred. No. 7.3e-239;
Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1  MHHHHHTAASDNFQSLGSGGQFAIPQGMATAGQIRSGGSPVTHIGPTAFGLGVVD  60
Db      1  MHHHHHTAASDNFQSLGSGGQFAIPQGMATAGQIRSGGSPVTHIGPTAFGLGVVD  60
Qy      61  NNGNGARVQRVGSAPASIGISTGDIVITAVDGAPINSATAMADALNGHPHGPDISVTWQ  120

```

Db 61 NNGCARVQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDIIVSTWQ 120
QY 121 TKSQGTTRTGNVTLAGGPPAEPMVDFGALPPEINARMYAGPGSASLVAAQMWDSVASDL 180
Db 121 TKSQGTTRTGNVTLAGGPPAEPMVDFGALPPEINARMYAGPGSASLVAAQMWDSVASDL 180
QY 181 PSAASAFQSVVWGLTVGSSWIGSSAGLMVAASPYVAMSVTAGOELTAAQVRAAAAYE 240
Db 181 PSAASAFQSVVWGLTVGSSWIGSSAGLMVAASPYVAMSVTAGOELTAAQVRAAAAYE 240
QY 241 TAYGLTVPPPIAENRAELMILATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPPIAENRAELMILATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPPEEAPMTSAGLLEQAAAEEASDTAAANQLMNNVPCALQQLAQTQGIT 360
Db 301 ATATATLLPPEEAPMTSAGLLEQAAAEEASDTAAANQLMNNVPCALQQLAQTQGIT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMVSMMNNHSMNTNSGVSMNTLSSMLKGFAPAPAAAQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNMVSMMNNHSMNTNSGVSMNTLSSMLKGFAPAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLSGSSLSGSGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLSGSSLSGSGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALP 480
QY 481 LTSLSAERPGQMLGGLPVGQMGARAGGGLSGVLVPRPYVMHSPHSPAGDIAPPALS 540
Db 481 LTSLSAERPGQMLGGLPVGQMGARAGGGLSGVLVPRPYVMHSPHSPAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVQVNVNINIKLGVNNAVAGTGIVDPNGVLTNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGVQVNVNINIKLGVNNAVAGTGIVDPNGVLTNNHVA 600
QY 601 GATDINAFSGSGQTYGVVVDYDRTQDVAVLQIRGAGGLPSAAIGGVAVGSPVVMGN 660
Db 601 GATDINAFSGSGQTYGVVVDYDRTQDVAVLQIRGAGGLPSAAIGGVAVGSPVVMGN 660
QY 661 SGQGGTPRAVPRGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGSPVVMGNLQ 720
Db 661 SGQGGTPRAVPRGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGSPVVMGNLQ 720
QY 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

RESULT 9

US-10-369-983-12
; Sequence 12, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guiderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 12
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: R95F (MTB72F-MAPS)
US-10-369-983-12

Query Match 99.9%; Score 3682; DB 15; Length 930;
Best Local Similarity 99.9%; Pred. No. 7,3e-239;
Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFILGLVVD 60
Db 1 MHHHHHTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFILGLVVD 60
QY 61 NNGCARVQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDIIVSTWQ 120
Db 61 NNGCARVQRVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDIIVSTWQ 120
QY 121 TKSQGTTRTGNVTLAGGPPAEPMVDFGALPPEINARMYAGPGSASLVAAQMWDSVASDL 180
Db 121 TKSQGTTRTGNVTLAGGPPAEPMVDFGALPPEINARMYAGPGSASLVAAQMWDSVASDL 180
QY 181 PSAASAFQSVVWGLTVGSSWIGSSAGLMVAASPYVAMSVTAGOELTAAQVRAAAAYE 240
Db 181 PSAASAFQSVVWGLTVGSSWIGSSAGLMVAASPYVAMSVTAGOELTAAQVRAAAAYE 240
QY 241 TAYGLTVPPPIAENRAELMILATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPPIAENRAELMILATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPPEEAPMTSAGLLEQAAAEEASDTAAANQLMNNVPCALQQLAQTQGIT 360
Db 301 ATATATLLPPEEAPMTSAGLLEQAAAEEASDTAAANQLMNNVPCALQQLAQTQGIT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMVSMMNNHSMNTNSGVSMNTLSSMLKGFAPAPAAAQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNMVSMMNNHSMNTNSGVSMNTLSSMLKGFAPAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLSGSSLSGSGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLSGSSLSGSGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALP 480
QY 481 LTSLSAERPGQMLGGLPVGQMGARAGGGLSGVLVPRPYVMHSPHSPAGDIAPPALS 540
Db 481 LTSLSAERPGQMLGGLPVGQMGARAGGGLSGVLVPRPYVMHSPHSPAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVQVNVNINIKLGVNNAVAGTGIVDPNGVLTNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGVQVNVNINIKLGVNNAVAGTGIVDPNGVLTNNHVA 600
QY 601 GATDINAFSGSGQTYGVVVDYDRTQDVAVLQIRGAGGLPSAAIGGVAVGSPVVMGN 660
Db 601 GATDINAFSGSGQTYGVVVDYDRTQDVAVLQIRGAGGLPSAAIGGVAVGSPVVMGN 660
QY 661 SGQGGTPRAVPRGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGSPVVMGNLQ 720
Db 661 SGQGGTPRAVPRGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGSPVVMGNLQ 720
QY 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

RESULT 10

US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guiderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85b)
US-10-369-983-18

Query Match          99.9%; Score 3682; DB 15; Length 1016;
Best Local Similarity 99.9%; Pred. No. 8.2e-239;
Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVD 60
DB 1 MHHHHHTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVD 60
QY 61 NNGNGARVQVRVGSAPAAASLGISTGVDVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
DB 61 NNGNGARVQVRVGSAPAAASLGISTGVDVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
QY 121 TKSGGTRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDL 180
DB 121 TKSGGTRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGSAGLWMAAASPYVAMSVTAGQBELTAAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSGSAGLWMAAASPYVAMSVTAGQBELTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPPPVIAENRAELMILIAIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPPPVIAENRAELMILIAIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPPEAPEMTSAGLLEQAAAVEASDTAAANQLMNNVPOALQOLAQPTQGT 360
DB 301 ATATATLLPPEAPEMTSAGLLEQAAAVEASDTAAANQLMNNVPOALQOLAQPTQGT 360
QY 361 PSSKLGGLWKTVSPHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTVSPHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
QY 481 LTSLTSAAEERPGQMLGGLPVGMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
DB 481 LTSLTSAAEERPGQMLGGLPVGMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVIDPENGVLVLTNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVIDPENGVLVLTNNHVA 600
QY 601 GATDINAFSVSGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAGGGVAVGEPVVMGN 660
DB 601 GATDINAFSVSGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAGGGVAVGEPVVMGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDDAIIQFGDAGGPPVNGLQ 720
DB 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDDAIIQFGDAGGPPVNGLQ 720
QY 721 VVGMTAAS 729
DB 721 VVGMTAAS 729
```

RESULT 11

US-10-369-983-17

; Sequence 17, Application US/10369983

; Publication No. US20030235593A1

; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir

```
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1022
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB102tm2F (MTB102FTM2, MTB72F-HTCC#1)
US-10-369-983-17
```

```
Query Match          99.9%; Score 3682; DB 15; Length 1022;
Best Local Similarity 99.9%; Pred. No. 8.3e-239;
Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVD 60
DB 1 MHHHHHTAASDNFQLSQGGGQFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVD 60
QY 61 NNGNGARVQVRVGSAPAAASLGISTGVDVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
DB 61 NNGNGARVQVRVGSAPAAASLGISTGVDVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
QY 121 TKSGGTRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDL 180
DB 121 TKSGGTRTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGSAGLWMAAASPYVAMSVTAGQBELTAAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSGSAGLWMAAASPYVAMSVTAGQBELTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPPPVIAENRAELMILIAIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPPPVIAENRAELMILIAIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPPEAPEMTSAGLLEQAAAVEASDTAAANQLMNNVPOALQOLAQPTQGT 360
DB 301 ATATATLLPPEAPEMTSAGLLEQAAAVEASDTAAANQLMNNVPOALQOLAQPTQGT 360
QY 361 PSSKLGGLWKTVSPHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
DB 361 PSSKLGGLWKTVSPHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
QY 421 TAAQNGVRAMSSLSGSLGSGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLSGSLGSGGVAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
QY 481 LTSLTSAAEERPGQMLGGLPVGMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
DB 481 LTSLTSAAEERPGQMLGGLPVGMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVIDPENGVLVLTNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVIDPENGVLVLTNNHVA 600
QY 601 GATDINAFSVSGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAGGGVAVGEPVVMGN 660
DB 601 GATDINAFSVSGQTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAGGGVAVGEPVVMGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDDAIIQFGDAGGPPVNGLQ 720
DB 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDDAIIQFGDAGGPPVNGLQ 720
QY 721 VVGMTAAS 729
```

```
Db 721 VVGNMNTAAS 729
|||||
|
RESULT 12
US-10-369-983-16
; Sequence 16, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein
; OTHER INFORMATION: MTB114F (MTB72P-mtCC#2)
US-10-369-983-16

Query Match 99.9%; Score 3682; DB 15; Length 1154;
Best Local Similarity 99.9%; Pred. No. 9,7e-239;
Matches 728; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAWAIGAIRSGGSPVTHIGTAFILGIVVD 60
DB 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAWAIGAIRSGGSPVTHIGTAFILGIVVD 60
QY 61 NNGNGARVQRVVGSAAPASIGISTGVDITAVDGAPINSATAMADALNGHHPGDI SVTWQ 120
DB 61 NNGNGARVQRVVGSAAPASIGISTGVDITAVDGAPINSATAMADALNGHHPGDI SVTWQ 120
QY 121 TKSGETGTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
DB 121 TKSGETGTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWTSAGSLMVAASPYVAMSVTAQCAELTAAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSGWTSAGSLMVAASPYVAMSVTAQCAELTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVIAENRAELMILITATNLGONTPTAIAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPVIAENRAELMILITATNLGONTPTAIAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPPEEAPEMTSAGLLEQAAVEEASDTAAANQLMNVPAQLQQAQPTQGT 360
DB 301 ATATATLLPPEEAPEMTSAGLLEQAAVEEASDTAAANQLMNVPAQLQQAQPTQGT 360
QY 361 PSSKLGWLKWTVPVPHSPISNNYVMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQ 420
DB 361 PSSKLGWLKWTVPVPHSPISNNYVMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLSGSGGVAANLGRAASVGSLSVPOAWAANAQVTPPARALP 480
DB 421 TAAQNGVRAMSSLGSSLSGSGGVAANLGRAASVGSLSVPOAWAANAQVTPPARALP 480
QY 481 LTSLSAABRGPCQMLGGLPVGMQARAGGSLGVLRVPRPVMPHSPHSPHSPHSPHSPH 540
DB 481 LTSLSAABRGPCQMLGGLPVGMQARAGGSLGVLRVPRPVMPHSPHSPHSPHSPHSPH 540
QY 541 QDRFADFPPALPLDPSAMVAQGVQVNVNTKLGYNNAVAGGTGIVIDPNGVLTNNHVIA 600
DB 541 QDRFADFPPALPLDPSAMVAQGVQVNVNTKLGYNNAVAGGTGIVIDPNGVLTNNHVIA 600

QY 601 GATDINAFSVGSGQTYGVYDRTQDVAVLQLRGAGLPSAALGGVAVGEPVVMANGN 660
DB 601 GATDINAFSVGSGQTYGVYDRTQDVAVLQLRGAGLPSAALGGVAVGEPVVMANGN 660
QY 661 SGGQGGTTPRAVPRVVALGQTVQASDSLTGAETLNGLIQFDDAATQPGDAGGPPVNGLGQ 720
DB 661 SGGQGGTTPRAVPRVVALGQTVQASDSLTGAETLNGLIQFDDAATQPGDAGGPPVNGLGQ 720
QY 721 VVGNMNTAAS 729
DB 721 VVGNMNTAAS 729

RESULT 13
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-2

Query Match 99.8%; Score 3677; DB 9; Length 729;
Best Local Similarity 99.7%; Pred. No. 1.2e-238;
Matches 727; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAWAIGAIRSGGSPVTHIGTAFILGIVVD 60
DB 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAWAIGAIRSGGSPVTHIGTAFILGIVVD 60
QY 61 NNGNGARVQRVVGSAAPASIGISTGVDITAVDGAPINSATAMADALNGHHPGDI SVTWQ 120
DB 61 NNGNGARVQRVVGSAAPASIGISTGVDITAVDGAPINSATAMADALNGHHPGDI SVTWQ 120
QY 121 TKSGETGTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
DB 121 TKSGETGTGNVTLAEGPPAEFMDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWTSAGSLMVAASPYVAMSVTAQCAELTAAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSGWTSAGSLMVAASPYVAMSVTAQCAELTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVIAENRAELMILITATNLGONTPTAIAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPVIAENRAELMILITATNLGONTPTAIAVNEAEYGEWMAQDAAMFGYAAAT 300
```

QY 301 ATATATLLPFEAPEMTSAGGELLEQAAAVEASDTAAANQLMNNVPOALQLOAQTQGT 360
 DB 301 ATATATLLPFEAPEMTSAGGELLEQAAAVEASDTAAANQLMNNVPOALQLOAQTQGT 360
 QY 361 PSSKLGGLWKTVPSPHRSPTSNMVSNNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
 DB 361 PSSKLGGLWKTVPSPHRSPTSNMVSNNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
 QY 421 TAAQNGVRAMSSLSGSSGLGGGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALP 480
 DB 421 TAAQNGVRAMSSLSGSSGLGGGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALP 480
 QY 481 LTSLSAAERGPQMLGGLPVQOMGARAGGSLGVLVPPRYPMHSPFAAGDIAPPALS 540
 DB 481 LTSLSAAERGPQMLGGLPVQOMGARAGGSLGVLVPPRYPMHSPFAAGDIAPPALS 540
 QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINIKLGYNNVAGAGTGIVIDPNGVVLNNHVA 600
 DB 541 QDRFADFPALPLDPSAMVAQVGPQVNNINIKLGYNNVAGAGTGIVIDPNGVVLNNHVA 600
 QY 601 GATDINAFSVSGQTYGVDVYDRTQDVAVLQLRGAGGLPSAAIQPGDAGGVPVWAMGN 660
 DB 601 GATDINAFSVSGQTYGVDVYDRTQDVAVLQLRGAGGLPSAAIQPGDAGGVPVWAMGN 660
 QY 661 SGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDAGGVPVWAMGN 720
 DB 661 SGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDAGGVPVWAMGN 720
 QY 721 VVGMNTAAS 729
 DB 721 VVGMNTAAS 729

RESULT 14
 US-09-886-349A-16
 ; Sequence 16, Application US/09886349A
 ; Publication No. US20040086523A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009070US
 ; CURRENT APPLICATION NUMBER: US/09/886, 349A
 ; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: US 09/597,796
 ; PRIOR FILING DATE: 2000-06-20
 ; PRIOR APPLICATION NUMBER: US 60/255,737
 ; PRIOR FILING DATE: 2001-02-01
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 16
 ; LENGTH: 729
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: tti-fusion
 ; OTHER INFORMATION: Protein MTB72F (Ra12-TbH9-Ra35 or MTB32-MTB39
 ; OTHER INFORMATION: Fusion)
 ; US-09-886-349A-16

Query Match 99.8%; Score 3677; DB 12; Length 729;
 Best Local Similarity 99.7%; Pred. No. 1.2e-238;
 Matches 727; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDFQLSQGGGFAIPGQAMATAGQIRSGGSPYHIGTAFGLGVVD 60
 DB 1 MHHHHHTAASDFQLSQGGGFAIPGQAMATAGQIRSGGSPYHIGTAFGLGVVD 60
 QY 61 NNGNGARVQVVGSAASLGISTGDIVTAVDGPINSATAMADALNHHHPGDIVSVTWQ 120

DB 61 NNGNGARVQVVGSAASLGISTGDIVTAVDGPINSATAMADALNHHHPGDIVSVTWQ 120
 QY 121 TKSGGTRTGNVTLLAEGPPAEPMVDFGALPEEINSARWVAGPGSASLVAAAQMDSDVASDL 180
 DB 121 TKSGGTRTGNVTLLAEGPPAEPMVDFGALPEEINSARWVAGPGSASLVAAAQMDSDVASDL 180
 QY 181 FSAASAFQSVVWGLTVGSSAGLMVAASAPYVAMSVTAQAEELTAQVRAAAAYE 240
 DB 181 FSAASAFQSVVWGLTVGSSAGLMVAASAPYVAMSVTAQAEELTAQVRAAAAYE 240
 QY 241 TAYGLTVPPVPIAENRAELMILITATNLLQNTTALIAVNEAEYGEWMAQDAAMFGYAAAT 300
 DB 241 TAYGLTVPPVPIAENRAELMILITATNLLQNTTALIAVNEAEYGEWMAQDAAMFGYAAAT 300
 QY 301 ATATATLLPFEAPEMTSAGGELLEQAAAVEASDTAAANQLMNNVPOALQLOAQTQGT 360
 DB 301 ATATATLLPFEAPEMTSAGGELLEQAAAVEASDTAAANQLMNNVPOALQLOAQTQGT 360
 QY 361 PSSKLGGLWKTVPSPHRSPTSNMVSNNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
 DB 361 PSSKLGGLWKTVPSPHRSPTSNMVSNNHMTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
 QY 421 TAAQNGVRAMSSLSGSSGLGGGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALP 480
 DB 421 TAAQNGVRAMSSLSGSSGLGGGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALP 480
 QY 481 LTSLSAAERGPQMLGGLPVQOMGARAGGSLGVLVPPRYPMHSPFAAGDIAPPALS 540
 DB 481 LTSLSAAERGPQMLGGLPVQOMGARAGGSLGVLVPPRYPMHSPFAAGDIAPPALS 540
 QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINIKLGYNNVAGAGTGIVIDPNGVVLNNHVA 600
 DB 541 QDRFADFPALPLDPSAMVAQVGPQVNNINIKLGYNNVAGAGTGIVIDPNGVVLNNHVA 600
 QY 601 GATDINAFSVSGQTYGVDVYDRTQDVAVLQLRGAGGLPSAAIQGGVAVGEPVWAMGN 660
 DB 601 GATDINAFSVSGQTYGVDVYDRTQDVAVLQLRGAGGLPSAAIQGGVAVGEPVWAMGN 660
 QY 661 SGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDAGGVPVWAMGN 720
 DB 661 SGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDAGGVPVWAMGN 720
 QY 721 VVGMNTAAS 729
 DB 721 VVGMNTAAS 729

RESULT 15
 US-10-359-460-2
 ; Sequence 2, Application US/10359460
 ; Publication No. US20030147911A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
 ; TITLE OF INVENTION: and Their Uses
 ; FILE REFERENCE: 014058-009020US
 ; CURRENT APPLICATION NUMBER: US/10/359,460
 ; CURRENT FILING DATE: 2003-02-05
 ; PRIOR APPLICATION NUMBER: US/09/287,849
 ; PRIOR FILING DATE: 1999-04-07
 ; PRIOR APPLICATION NUMBER: US 08/818,112
 ; PRIOR FILING DATE: 1997-03-13
 ; PRIOR APPLICATION NUMBER: US 08/942,578
 ; PRIOR FILING DATE: 1997-10-01
 ; PRIOR APPLICATION NUMBER: US 09/025,197
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 09/056,556
 ; PRIOR FILING DATE: 1998-04-07

; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-10-359-460-2

Query Match 99.8%; Score 3677; DB 14; Length 729;
Best Local Similarity 99.7%; Pred. No. 1.2e-238;
Matches 727; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MHHHHHTAASDNFQLSGGCGGPAIPICQAWAIAGQIRSGGSGPTVHIGTAFGLGVVD	60
Db	1	MHHHHHTAASDNFQLSGGCGGPAIPICQAWAIAGQIRSGGSGPTVHIGTAFGLGVVD	60
Qy	61	NNNGARVQRVVGSAPASLGISTGDIVITAVDGAPINSATAMADALNGHHPGDVISVTWQ	120
Db	61	NNNGARVQRVVGSAPASLGISTGDIVITAVDGAPINSATAMADALNGHHPGDVISVTWQ	120
Qy	121	TKSGGTRTGNVTLAEGPPAPFMDVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL	180
Db	121	TKSGGTRTGNVTLAEGPPAPFMDVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL	180
Qy	181	FSAASAFQSVVWGLTVGSGWGLVVAASPVVAMSVTAGQAEITAAQVEVAAAAYE	240
Db	181	FSAASAFQSVVWGLTVGSGWGLVVAASPVVAMSVTAGQAEITAAQVEVAAAAYE	240
Qy	241	TAYGLTVPPVPIAENRAELMILIAITNLGQNTFAIAVNEAEYGEWMAQDAAMFGYAAAT	300
Db	241	TAYGLTVPPVPIAENRAELMILIAITNLGQNTFAIAVNEAEYGEWMAQDAAMFGYAAAT	300
Qy	301	ATATATLLPPEEAPEMTSAGLLFQAAVVEASDTPAAANQLMNNVPAALQQAQPTQGT	360
Db	301	ATATATLLPPEEAPEMTSAGLLFQAAVVEASDTPAAANQLMNNVPAALQQAQPTQGT	360
Qy	361	PSSKLGGLKWTSPHRSPISNMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQ	420
Db	361	PSSKLGGLKWTSPHRSPISNMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQ	420
Qy	421	TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOAANAQAVTPAARALP	480
Db	421	TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOAANAQAVTPAARALP	480
Qy	481	LTSLSAAERGPQMLGGLPVGQMGARAGGLSGVLRVPPRPVPMPSAAGDIAPPALS	540
Db	481	LTSLSAAERGPQMLGGLPVGQMGARAGGLSGVLRVPPRPVPMPSAAGDIAPPALS	540
Qy	541	QDRFADFPALPLDPSAVQAQVGVQVNNIKLGYNNAVGAGTIVIDPNGVLTNNHVA	600
Db	541	QDRFADFPALPLDPSAVQAQVGVQVNNIKLGYNNAVGAGTIVIDPNGVLTNNHVA	600
Qy	601	GATDINAFSVGSGQTGVGVVVDRTQDVAVLQRCAGGIPSAATGGGVAVGEPVVMGN	660
Db	601	GATDINAFSVGSGQTGVGVVVDRTQDVAVLQRCAGGIPSAATGGGVAVGEPVVMGN	660
Qy	661	SGCGGTTPRAVGRVVALGQTQVQASDLTCAETLNGLIQFDAAIQPDAGGPPVWNLGQ	720
Db	661	SGCGGTTPRAVGRVVALGQTQVQASDLTCAETLNGLIQFDAAIQPDAGGPPVWNLGQ	720
Qy	721	VGMNTAAS 729	
Db	721	VGMNTAAS 729	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:07:54 ; Search time 21.2754 Seconds
(without alignments)
1446.225 Million cell updates/sec

Title: US-09-886-349A-20

Perfect score: 3007

Sequence: 1 HHHHHHHHVDGALPPEIN.....SGGPVNGLGQVGMNTAAS 596

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/protdata/2/aaa/5A_COMB.pep:*
2: /cgn2_6/protdata/2/aaa/5B_COMB.pep:*
3: /cgn2_6/protdata/2/aaa/5A_COMB.pep:*
4: /cgn2_6/protdata/2/aaa/5B_COMB.pep:*
5: /cgn2_6/protdata/2/aaa/PTUS_COMB.pep:*
6: /cgn2_6/protdata/2/aaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3007	100.0	596	4	US-09-287-849-26
2	2941	97.8	729	4	US-09-223-040-2
3	2941	97.8	729	4	US-09-287-849-2
4	2033.5	67.6	600	4	US-09-287-849-22
5	1949	64.8	391	3	US-08-818-112-107
6	1949	64.8	391	4	US-08-818-111-102
7	1949	64.8	391	4	US-09-056-556-107
8	1949	64.8	391	4	US-09-072-967-102
9	1949	64.8	391	4	US-09-072-967-107
10	1652.5	55.0	396	3	US-08-818-112-111
11	1652.5	55.0	396	4	US-08-818-111-106
12	1652.5	55.0	396	4	US-09-056-556-111
13	1652.5	55.0	396	4	US-09-072-967-106
14	1652.5	55.0	396	4	US-09-072-967-111
15	1486.5	49.4	359	3	US-08-818-111-109
16	1486.5	49.4	359	4	US-08-818-111-104
17	1486.5	49.4	359	4	US-09-056-556-109
18	1486.5	49.4	359	4	US-09-072-967-104
19	1486.5	49.4	359	4	US-08-818-112-91
20	1189	39.5	263	3	US-08-818-112-91
21	1187	39.5	263	4	US-08-818-111-92
22	1187	39.5	263	4	US-09-056-556-91
23	1187	39.5	263	4	US-09-072-967-92
24	1187	39.5	263	4	US-09-072-967-91
25	1187	39.5	263	4	US-08-818-112-79
26	983	32.7	355	3	US-08-818-112-79
27	983	32.7	355	4	US-08-818-111-80

28	983	32.7	355	4	US-09-056-556-79	Sequence 79, Appl
29	983	32.7	355	4	US-09-072-967-80	Sequence 80, Appl
30	983	32.7	355	4	US-09-072-967-79	Sequence 79, Appl
31	770.5	25.6	400	4	US-09-073-009-126	Sequence 126, Appl
32	770.5	25.6	400	4	US-09-073-010-126	Sequence 126, Appl
33	607.5	20.2	710	4	US-09-287-849-16	Sequence 16, Appl
34	607.5	20.2	856	4	US-09-287-849-12	Sequence 12, Appl
35	606.5	20.2	423	4	US-08-073-009-142	Sequence 142, Appl
36	606.5	20.2	423	4	US-09-073-010-142	Sequence 142, Appl
37	450.5	15.0	943	4	US-09-477-135A-131	Sequence 131, Appl
38	424	14.1	141	4	US-09-073-009-15	Sequence 15, Appl
39	424	14.1	141	4	US-09-073-010-15	Sequence 15, Appl
40	381.5	12.7	204	4	US-08-311-731A-57	Sequence 57, Appl
41	377.5	12.6	208	4	US-08-311-731A-208	Sequence 208, Appl
42	328.5	10.9	371	4	US-09-050-739-92	Sequence 92, Appl
43	325.5	10.8	368	3	US-08-818-112-114	Sequence 114, Appl
44	325.5	10.8	368	4	US-08-818-111-109	Sequence 109, Appl
45	325.5	10.8	368	4	US-09-056-556-114	Sequence 114, Appl

ALIGNMENTS

RESULT 1
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 100.0%; Score 3007; DB 4; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.3e-219;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 HHHHHHHHVDGALPPEINSGGPVNGLGQVGMNTAASAFQSVVWG 60
Db 1 HHHHHHHHVDGALPPEINSGGPVNGLGQVGMNTAASAFQSVVWG 60
QY 61 LTGVSWIGSSAGLVAAAPVYVNMSTAGQELTAQVRVAAAAYETAYGLTVPFPPVIA 120
Db 61 LTGVSWIGSSAGLVAAAPVYVNMSTAGQELTAQVRVAAAAYETAYGLTVPFPPVIA 120
QY 121 ENRAELMILLIATNLLGQNTPAIVNNEARYGENWQAADAAAFGAAATATATATLPPFEA 180

Db 121 ENRAELMILIAITNLGQNTPAIAVNEAYEGMWAQDAAMFGYAAATATATATALLPPEEA 180
QY 181 PEMTSAGLLBQAAAVEASDTAAANQLMNNVPAALQQAQPTQGTTPSSKLGGLWKTIVS 240
Db 181 PEMTSAGLLBQAAAVEASDTAAANQLMNNVPAALQQAQPTQGTTPSSKLGGLWKTIVS 240
QY 241 PHRSPISNMWMMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSL 300
Db 241 PHRSPISNMWMMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSL 300
QY 301 GSSLGSSGLGGVAANIIGRAASVGSLSVPQAWAANAQAVTPAARALPLTSLTSAASRGPG 360
Db 301 GSSLGSSGLGGVAANIIGRAASVGSLSVPQAWAANAQAVTPAARALPLTSLTSAASRGPG 360
QY 361 QMLGGLPVQMGARAGGGLSVLRVPPRYMPHSPAAAGDIAPPALSQDRFADFPALPLD 420
Db 361 QMLGGLPVQMGARAGGGLSVLRVPPRYMPHSPAAAGDIAPPALSQDRFADFPALPLD 420
QY 421 PSAMVAQVGPQVNNINTKLYNNNAVAGTGIIVDPNGVLTNNHVIAGATDINAFSVGSG 480
Db 421 PSAMVAQVGPQVNNINTKLYNNNAVAGTGIIVDPNGVLTNNHVIAGATDINAFSVGSG 480
QY 481 QTYGVDVYDRTQDVAVLQLRGAGGLPSAAIIGGVAVGEPVVMGNSGGQGTTPRAVPG 540
Db 481 QTYGVDVYDRTQDVAVLQLRGAGGLPSAAIIGGVAVGEPVVMGNSGGQGTTPRAVPG 540
QY 541 RVVALGQTVQASDLSLTGAETTLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAAS 596
Db 541 RVVALGQTVQASDLSLTGAETTLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAAS 596

RESULT 2
US-09-223-040-2
; Sequence 2, Application US/09223040
; Patent No. 6544522
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-223-040-2

Query Match 97.8%; Score 2941; DB 4; Length 729;
Best Local Similarity 99.8%; Pred. No. 1.7e-214;
Matches 587; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 MVDFFGALPPEINSARMYAGPGSASLVAAQMDVSDVLSAASAFQSVVWGLTVGSMIG 68
Db 142 MVDFFGALPPEINSARMYAGPGSASLVAAQMDVSDVLSAASAFQSVVWGLTVGSMIG 201

QY 69 SSAGLMVAASPYVAMSVTAGAELTAAQVRAAAAYETAYGLTVPPPPVIAENRAELMI 128
Db 202 SSAGLMVAASPYVAMSVTAGAELTAAQVRAAAAYETAYGLTVPPPPVIAENRAELMI 261

QY 129 LIATNLGQNTPAIAVNEAYEGMWAQDAAMFGYAAATATATATALLPPEAPEMTSAGG 188
Db 262 LIATNLGQNTPAIAVNEAYEGMWAQDAAMFGYAAATATATATALLPPEAPEMTSAGG 321

QY 189 LLEQAAAVEASDTAAANQLMNNVPAALQQAQPTQGTTPSSKLGGLWKTIVSPHRSPISN 248

Db 322 LLEQAAAVEASDTAAANQLMNNVPAALQQAQPTQGTTPSSKLGGLWKTIVSPHRSPISN 381
QY 249 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 308
Db 382 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 441
QY 309 LGGGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALPLTSLTSAASRGPGQMLGGLPV 368
Db 442 LGGGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALPLTSLTSAASRGPGQMLGGLPV 501
QY 369 GQMGARAGGGLSVLRVPPRYMPHSPAAAGDIAPPALSQDRFADFPALPLDPSAMVAQV 428
Db 502 GQMGARAGGGLSVLRVPPRYMPHSPAAAGDIAPPALSQDRFADFPALPLDPSAMVAQV 561
QY 429 GPQVNNINTKLYNNNAVAGTGIIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDVV 488
Db 562 GPQVNNINTKLYNNNAVAGTGIIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDVV 621
QY 489 GYDRTQDVAVLQLRGAGGLPSAAIIGGVAVGEPVVMGNSGGQGTTPRAVPGVVALGQT 548
Db 622 GYDRTQDVAVLQLRGAGGLPSAAIIGGVAVGEPVVMGNSGGQGTTPRAVPGVVALGQT 681
QY 549 VOASDLSLTGAETTLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAAS 596
Db 682 VOASDLSLTGAETTLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAAS 729

RESULT 3
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-2

Query Match 97.8%; Score 2941; DB 4; Length 729;
Best Local Similarity 99.8%; Pred. No. 1.7e-214;
Matches 587; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 9 MVDFFGALPPEINSARMYAGPGSASLVAAQMDVSDVLSAASAFQSVVWGLTVGSMIG 68
Db 142 MVDFFGALPPEINSARMYAGPGSASLVAAQMDVSDVLSAASAFQSVVWGLTVGSMIG 201

QY 69 SSAGLMVAASPYVAMSVTAGAELTAAQVRAAAAYETAYGLTVPPPPVIAENRAELMI 128

Db 202 SSAGLWAAAGPYVAVMSVTAQAEILTAAQVRVAAAAVETAYGLTVPVPVIAENRAELMI 261
Qy 129 LIATNLLGONTPTAIIVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPENTSGG 188
Db 262 LIATNLLGONTPTAIIVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPENTSGG 321
Qy 189 LLEQAAAEEASDPTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTYSPHRSPISN 248
Db 322 LLEQAAAEEASDPTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTYSPHRSPISN 381
Qy 249 MYSMANNHMTNSGVSWTNTLSMLKGFAPAAAAQAQVCTAAQNGVRAMSSLSGSSG 308
Db 382 MYSMANNHMTNSGVSWTNTLSMLKGFAPAAAAQAQVCTAAQNGVRAMSSLSGSSG 441
Qy 309 LGGGVAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAARPGQMLGGLPV 368
Db 442 LGGGVAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAARPGQMLGGLPV 501
Qy 369 GQMGARAGGGLSGVLVRPVPYVMPHSPAAGDIAPPALSQRRFADFPALPDPSAMVAQV 428
Db 502 GQMGARAGGGLSGVLVRPVPYVMPHSPAAGDIAPPALSQRRFADFPALPDPSAMVAQV 561
Qy 429 GPQVNVNITKLYNNNAVAGTGTGIVDPNGVLTNNHVIAGATDINAFSVGSQTYGVVDV 488
Db 562 GPQVNVNITKLYNNNAVAGTGTGIVDPNGVLTNNHVIAGATDINAFSVGSQTYGVVDV 621
Qy 489 GYDRTQDVAVLQLRGAGGLPSAAIGGVAVGVPVAVMGNSGGQGTTPRAVGRVVALQOT 548
Db 622 GYDRTQDVAVLQLRGAGGLPSAAIGGVAVGVPVAVMGNSGGQGTTPRAVGRVVALQOT 681
Qy 549 VQASDLSLTGABETLNGLIQFDAAIQGDSGGVWNLGOVGMNTAAS 596
Db 682 VQASDLSLTGABETLNGLIQFDAAIQGDSGGVWNLGOVGMNTAAS 729

RESULT 4

US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-005020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

Query Match 67.6%; Score 2033.5; DB 4; Length 600;
Best Local Similarity 72.5%; Pred. No. 6.7e-146;

Matches 440; Conservative 19; Mismatches 81; Indels 67; Gaps 10;
Qy 1 HMEHHHHHVVDFGALPEINSARMYAGPGSASILVAAAQWMDSVASDLFSAASAFQSVVWG 60
Db 1 HMEHHHHHVVDFGALPEINSARMYAGPGSASILVAAAQWMDSVASDLFSAASAFQSVVWG 60
Qy 61 LTVGSWTGSSAGLWMAAASPYPVAVMSVTAQAEILTAAQVRVAAAAVETAYGLTVPVPVIA 120
Db 61 LTVGSWTGSSAGLWMAAASPYPVAVMSVTAQAEILTAAQVRVAAAAVETAYGLTVPVPVIA 120
Qy 121 ENRAELMILATNLLGONTPTAIIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEA 180
Db 121 ENRAELMILATNLLGONTPTAIIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEA 180
Qy 181 PEMTSAGGLLEQAAAEEASDPTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTYS 240
Db 181 PEMTSAGGLLEQAAAEEASDPTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTYS 240
Qy 241 PHRSPISNMYSMANNHMTNSGVSWTNTLSMLKGFAPAAAAQAQVCTAAQNGVRAMSSSL 300
Db 241 PHRSPISNMYSMANNHMTNSGVSWTNTLSMLKGFAPAAAAQAQVCTAAQNGVRAMSSSL 300
Qy 301 GSSILGSSGLGGVVAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAARPG 360
Db 301 GSSILGSSGLGGVVAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAARPG 360
Qy 361 QMLGGLPVGQMGARAGGGLSGVLVRPVPYVMPHSPAAGDIAP 403
Db 361 QMLGGLPVGQMGARAGGGLSGVLVRPVPYVMPHSPAAGDIAP 420
Qy 404 -PALSQDRFADFPALPDPS 448
Db 421 LNATDPGAAQFNASPVASVLRNFLAAPPFPAAMAAQL-QAVFGAAQYIGLVESVAGS 479
Qy 449 TGIVIDPNGVLTNNHVIAGATDINAFSVGSQTYGVVDVGVYDRTQDVAVLQLRGAGGLP 508
Db 480 CN-----NYELMTINYQFG---DVDA-----HGAMIRAQAASLEAEHQAIVRDVLAAQDFW 527
Qy 509 SAAIGGGVAVGEPVAVMG-----NSGGQGTTPRAVGRVVALQOTVQASDLSLTGA 558
Db 528 GGA--GSVACEFITQLGRNFQVIYEQANAHGQ-----KYQAGNNNAQOTDSAVGS 576
Qy 559 E-ETLNG 564
Db 577 SWATSN 583

RESULT 5

US-08-818-112-107
; Sequence 107, Application US/08818112
; Patent No. 6290369
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818.112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-107

Query Match 64.8%; Score 1949; DB 3; Length 391;
Best Local Similarity 100.0%; Pred. No. 9.4e-140;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 MVDFGALPPPEINSGARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 68
DB 1 MVDFGALPPPEINSGARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 60
QY 69 SSAGLMVAASPPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 128
DB 61 SSAGLMVAASPPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
QY 129 LIATNLGQNTPAIVNAEYEGEMWQDAAMFGYAAATATATATLLPPEEAPMTSAGG 188
DB 121 LIATNLGQNTPAIVNAEYEGEMWQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180
QY 189 LLEQAAAVEASDTPAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 248
DB 181 LLEQAAAVEASDTPAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
QY 249 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSLGSSG 308
DB 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSLGSSG 300
QY 309 LGGGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 368
DB 301 LGGGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
QY 369 GQMGARAGGGLSGVLRVPPRPYVPHSPAAG 399
DB 361 GQMGARAGGGLSGVLRVPPRPYVPHSPAAG 391

RESULT 6
US-08-818-111-102
Sequence 102, Application US/08818111
Patent No. 6338852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818.111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-102

Query Match 64.8%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 9.4e-140;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 MVDFGALPPPEINSGARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 68
DB 1 MVDFGALPPPEINSGARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 60
QY 69 SSAGLMVAASPPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 128
DB 61 SSAGLMVAASPPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 120
QY 129 LIATNLGQNTPAIVNAEYEGEMWQDAAMFGYAAATATATATLLPPEEAPMTSAGG 188
DB 121 LIATNLGQNTPAIVNAEYEGEMWQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180
QY 189 LLEQAAAVEASDTPAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 248
DB 181 LLEQAAAVEASDTPAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
QY 249 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSLGSSG 308
DB 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSLGSSG 300
QY 309 LGGGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 368
DB 301 LGGGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
QY 369 GQMGARAGGGLSGVLRVPPRPYVPHSPAAG 399
DB 361 GQMGARAGGGLSGVLRVPPRPYVPHSPAAG 391

RESULT 7
US-09-056-556-107
Sequence 107, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington

```

; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-056-556-107

Query Match 64.8%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 9.4e-140;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MYDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLPFSAASAFQSVVWGLTVGSMIG 68
DB 1 MYDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLPFSAASAFQSVVWGLTVGSMIG 60
QY 69 SSAGLWMAAASPYYVAMWSVTAGQAELTAAQVRVAAAAAYETAYGLTVPPVIAENRAELMI 128
DB 61 SSAGLWMAAASPYYVAMWSVTAGQAELTAAQVRVAAAAAYETAYGLTVPPVIAENRAELMI 120
QY 129 LIATNLLQGNTPATAVNEAEYGEVWAQDAAMFGYAAATATATATLTPPEEAPEMTSAGG 188
DB 121 LIATNLLQGNTPATAVNEAEYGEVWAQDAAMFGYAAATATATATLTPPEEAPEMTSAGG 180
QY 189 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPISN 248
DB 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGGLWKTVPSPHRSPISN 240
QY 249 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAAQVQTAAQNGVRAMSSLGSSLGSSG 308
DB 241 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAAQVQTAAQNGVRAMSSLGSSLGSSG 300
QY 309 LGGVVAANI GRAASVGSLSVPOAAANAQVTPAARALPLTSLTSSAAERGPQGMGLGLPV 368
DB 301 LGGVVAANI GRAASVGSLSVPOAAANAQVTPAARALPLTSLTSSAAERGPQGMGLGLPV 360
QY 369 GQMGARAGGGLSGVLRVPPRPYPVMPHSPAAG 399
DB 361 GQMGARAGGGLSGVLRVPPRPYPVMPHSPAAG 391

RESULT 8
US-09-072-596-102
; Sequence 102, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.

```

APPLICANT: Dillon, Davin C.
 APPLICANT: Campos-Neto, Antonio
 APPLICANT: Houghton, Raymond
 APPLICANT: Vedvick, Thomas S.
 APPLICANT: Twardzik, Daniel R.
 APPLICANT: Lodes, Michael J.
 APPLICANT: Hendrickson, Ronald C.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
 NUMBER OF SEQUENCES: 355
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/072,967
 FILING DATE: 05-MAY-1998
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.411C9
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 107:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 391 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-09-072-967-107

Query Match 64.8%; Score 1949; DB 4; Length 391;
 Best Local Similarity 100.0%; Pred. No. 9,4e-140; Indels 0; Gaps 0;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 9 MVDGALPPEINSARMYAGSGASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
 DB 1 MVDGALPPEINSARMYAGSGASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
 QY 69 SSAGLMVAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 128
 DB 61 SSAGLMVAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
 QY 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEAPEMTSAGG 188
 DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEAPEMTSAGG 180
 QY 189 LLEQAAVEAEASDTAAANQLMNNVPQALQOQPTGTTTSSKLGGLWKTVSPHRSPI SN 248
 DB 181 LLEQAAVEAEASDTAAANQLMNNVPQALQOQPTGTTTSSKLGGLWKTVSPHRSPI SN 240
 QY 249 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAACAVQTAQNGVRAMSSIGSLGSSG 308
 DB 241 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAACAVQTAQNGVRAMSSIGSLGSSG 300
 QY 309 LGGGVAANLGRAASVGSLSVPOQAAAAQVTPAARALPLTSLTSAABERGPGQMLGGLPV 368
 DB 301 LGGGVAANLGRAASVGSLSVPOQAAAAQVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
 QY 369 GOMGARAGGLSLVLRVPRPYVMPHSFAAG 399
 DB 361 GOMGARAGGLSLVLRVPRPYVMPHSFAAG 391

RESULT 10
 US-08-818-112-111
 ; Sequence 111, Application US/08818112
 ; Patent No. 6290969
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 153
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/818,112
 ; FILING DATE: 13-MAR-1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.411C6
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 111:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 396 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; US-08-818-112-111

Query Match 55.0%; Score 1652.5; DB 3; Length 396;
 Best Local Similarity 84.9%; Pred. No. 2.7e-117; Indels 7; Gaps 3;
 Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
 QY 9 MVDGALPPEINSARMYAGSGASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
 DB 1 MVDGALPPEINSARMYAGSGASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
 QY 69 SSAGLMVAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 128
 DB 61 SSAGLMVAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
 QY 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 188
 DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEAPEMTSAGG 180
 QY 189 LLEQAAVEAEASDTAAANQLMNNVPQALQOQPTGTTTSSKLGGLWKTVSPHRSPI SN 248
 DB 181 LLEQAAVEAEIDTAANQLMNNVPQALQOQPTKSIWPFQDLSLWKAISPHLSPLSN 240
 QY 249 MYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAACAVQTAQNGVRAMSSIGSLGSSL 304
 DB 241 IVSMLNNHSMNTNSGVSMNTLSSMLKGFAPAAAACAVQTAQNGVRAMSSIGSLGSSL 299
 QY 305 GSSGLGGGVAANLGRAASVGSLSVPOQAAAAQVTPAARALPLTSLTSAABERGPGQMLG 364

Db 300 GSSGLGAGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAQAQAPGHMLG 359

Qy 365 GLPVQGMGARAG--GGLSGVLRVPPRPYPVMPHSPAAG 399

Db 360 GLPLGQLTNSGGFGGVSNALRMPRAVYVMPRVPAAG 396

RESULT 11

US-08-818-111-106

; Sequence 106, Application US/08818111

; Patent No. 6338852

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Campos-Neto, Antonia

; APPLICANT: Houghton, Raymond

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Twardzik, Daniel R.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS

; NUMBER OF SEQUENCES: 148

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/818,111

; FILING DATE: 13-MAR-1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.417C6

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 106:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 396 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

US-08-818-111-106

Query Match 55.0%; Score 1652.5; DB 4; Length 396;

Best Local Similarity 84.9%; Pred. No. 2.7e-117;

Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

Qy 9 MVDFGALPPEINSAARMYAGPGSASLVAAQAQWDSVADLFSAAAFQSVVWGLTVGWSWG 68

Db 1 VVDFGALPPEINSAARMYAGPGSASLVAAQAQWDSVADLFSAAAFQSVVWGLTVGWSWG 60

Qy 69 SSAGLMVAASPSYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 128

Db 61 SSAGLMVAASPSYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120

Qy 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGG 188

Db 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEDAPLIINPGG 180

Qy 189 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTGTTTSSKLGGLWKTVPSPHRSPISN 248

Db 181 LLEQAAVEEAIDTAANQLMNNVPQALQQAQPTKSIWPFQDLSLWKAISPLSPLN 240

Qy 249 MVSNNMHNMTNSGVSMWTNTLSSMLKGFAPAAAQAQVIAQNGVRAWSS-----LGSSSL 304

Db 241 IVSMLNNHVSMTNSGVSMASTLHSLMLKGFAP-AAACAVETAQAQNGVQAMSSSLGSLGSSSL 299

Qy 305 GSSGLGAGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABEGPGOMLG 364

Db 300 GSSGLGAGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAQAQAPGHMLG 359

Qy 365 GLPVQGMGARAG--GGLSGVLRVPPRPYPVMPHSPAAG 399

Db 360 GLPLGQLTNSGGFGGVSNALRMPRAVYVMPRVPAAG 396

RESULT 12

US-09-056-556-111

; Sequence 111, Application US/09056556

; Patent No. 6350456

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND TREATME

; NUMBER OF SEQUENCES: 241

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/056,556

; FILING DATE: 07-APR-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.457

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 111:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 396 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

US-09-056-556-111

Query Match 55.0%; Score 1652.5; DB 4; Length 396;

Best Local Similarity 84.9%; Pred. No. 2.7e-117;

Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

Qy 9 MVDFGALPPEINSAARMYAGPGSASLVAAQAQWDSVADLFSAAAFQSVVWGLTVGWSWG 68

Db 1 VVDFGALPPEINSAARMYAGPGSASLVAAQAQWDSVADLFSAAAFQSVVWGLTVGWSWG 60

Qy 69 SSAGLMVAASPSYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 128

Db 61 SSAGLMVAASPSYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120

Qy 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGG 188

Db 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEDAPLIINPGG 180

Qy 189 LLEQAAAVEASDTAAANQLMNNVPQALQQAQPTGTTTSSKLGGLWKTVPSPHRSPISN 248

Db 181 LLEQAAVEEAIDTAANQLMNNVPQALQQAQPTKSIWPFQDLSLWKAISPLSPLN 240


```
Db 1 VVDFGALPPEINSARMYAGPGSASLVAAAKWMDSVASDLFSAASAFQSVVWGLTTGWSWIG 60
Qy 69 SSAGLMVAASAPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 128
Db 61 SSAGLMVAASAPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
Qy 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFVAAAATATATATLIPFEAPEMTSAGG 188
Db 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFVAAAATATATATLIPFEAPEMTSAGG 180
Qy 189 LLEQAAVVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPLSN 248
Db 181 LLEQAAVVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPLSN 240
Qy 249 MVSMANNHSMNTSGVSMNTLSSMLKGFAPAAAAQAVCTAAQNGVRAMSS-----LGSSL 304
Db 241 IVSMLNHHVSMNTSGVSMNTLSSMLKGFAPAAAAQAVCTAAQNGVRAMSS-----LGSSL 299
Qy 305 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 364
Db 300 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 359
Qy 365 GLFVGQMGARAG--GGLSGVLVPRPPRYMHPSPAAG 399
Db 360 GLFVLGQNTSGGGFGVSNALRMPPRAYMPPRVPAAAG 396
```

RESULT 15

```
US-08-818-112-109
; Sequence 109, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-818-112-109
```

```
Query Match 49.4%; Score 1486.5; DB 3; Length 359;
Best Local Similarity 84.2%; Pred. No. 8.8e-105;
Matches 303; Conservative 16; Mismatches 36; Indels 5; Gaps 2;
Qy 9 MVDFFGALPPEINSARMYAGPGSASLVAAAKWMDSVASDLFSAASAFQSVVWGLTTGWSWIG 68
Db 1 VVDFGALPPEINSARMYAGPGSASLVAAAKWMDSVASDLFSAASAFQSVVWGLTTGWSWIG 60
Qy 69 SSAGLMVAASAPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 128
Db 61 SSAGLMVAASAPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPPVIAENRAELMT 120
Qy 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFVAAAATATATATLIPFEAPEMTSAGG 188
Db 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFVAAAATATATATLIPFEAPEMTSAGG 180
Qy 189 LLEQAAVVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPLSN 248
Db 181 LLEQAAVVEEASDTAAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPLSN 240
Qy 249 MVSMANNHSMNTSGVSMNTLSSMLKGFAPAAAAQAVCTAAQNGVRAMSS-----LGSSL 304
Db 241 VSSIANNHSMNTSGVSMNTLSSMLKGFAPAAAAQAVCTAAQNGVRAMSS-----LGSSL 299
Qy 305 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 364
Db 300 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABERGPGOMLG 359
```

Search completed: June 22, 2004, 17:27:36

Job time : 22.2754 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 55.2012 seconds
(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-20

Perfect score: 3007

Sequence: 1 HMMHHHHMDFGALPPBEIN.....SGGPVNLGQVGMNTAAS 596

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3007	100.0	596	9	US-09-287-849-26
2	3007	100.0	596	12	US-09-886-349A-20
3	3007	100.0	596	14	US-10-359-460-26
4	3007	100.0	596	14	US-10-098-732A-20
5	2946	98.0	729	15	US-10-369-983-21
6	2946	98.0	813	15	US-10-369-983-15
7	2946	98.0	825	15	US-10-369-983-14
8	2946	98.0	875	15	US-10-369-983-13
9	2946	98.0	930	14	US-10-098-732A-65
10	2946	98.0	930	15	US-10-369-983-12
11	2946	98.0	1016	15	US-10-369-983-18
12	2946	98.0	1022	15	US-10-369-983-17
13	2946	98.0	1154	15	US-10-369-983-16
14	2943	97.9	729	15	US-09-886-349A-18
15	2943	97.9	729	14	US-10-098-732A-18

16	2943	97.9	729	15	US-10-369-983-22	Sequence 22, Appl
17	2941	97.8	729	9	US-09-287-849-2	Sequence 2, Appl
18	2941	97.8	729	12	US-09-886-349A-16	Sequence 16, Appl
19	2941	97.8	729	14	US-10-359-460-2	Sequence 2, Appl
20	2941	97.8	729	14	US-10-098-732A-16	Sequence 16, Appl
21	2941	97.8	729	15	US-10-359-459-22	Sequence 2, Appl
22	2033.5	67.6	600	9	US-09-287-849-22	Sequence 22, Appl
23	2033.5	67.6	600	14	US-10-359-460-22	Sequence 22, Appl
24	1955	65.0	1010	15	US-10-369-983-4	Sequence 4, Appl
25	1949	64.8	391	12	US-09-886-349A-14	Sequence 14, Appl
26	1949	64.8	391	14	US-10-193-002-102	Sequence 102, App
27	1949	64.8	391	14	US-10-084-843-107	Sequence 107, App
28	1949	64.8	391	14	US-10-098-732A-14	Sequence 14, Appl
29	1949	64.8	723	15	US-10-369-983-2	Sequence 8, Appl
30	1934	64.3	391	12	US-09-872-186-8	Sequence 106, App
31	1652.5	55.0	396	14	US-10-193-002-106	Sequence 111, App
32	1652.5	55.0	396	14	US-10-084-843-111	Sequence 62455, A
33	1583	52.6	393	12	US-10-282-122A-62455	Sequence 62455, A
34	1486.5	49.4	359	12	US-10-282-122A-64892	Sequence 104, App
35	1486.5	49.4	359	14	US-10-193-002-104	Sequence 104, App
36	1486.5	49.4	359	14	US-10-084-843-109	Sequence 8, Appl
37	1189	39.5	358	9	US-09-287-849-8	Sequence 8, Appl
38	1189	39.5	358	14	US-10-359-460-8	Sequence 12, Appl
39	1187	39.5	263	12	US-09-886-349A-12	Sequence 92, Appl
40	1187	39.5	263	14	US-10-193-002-92	Sequence 91, Appl
41	1187	39.5	263	14	US-10-084-843-91	Sequence 12, Appl
42	1187	39.5	263	14	US-10-098-732A-12	Sequence 161, App
43	989	32.9	355	9	US-09-712-363-161	Sequence 4, Appl
44	987	32.8	330	12	US-09-886-349A-4	Sequence 4, Appl
45	987	32.8	330	14	US-10-098-732A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-09-287-849-26 Application US/09287849
; Sequence 26 Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 100.0%; Score 3007; DB 9; Length 596;
Best Local Similarity 100.0%; Pred. No. 5e-202;

Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HMEHHHHMVDGALPPEINSMRYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWG 60
DB 1 HMEHHHHMVDGALPPEINSMRYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWG 60

QY 61 LTVGSWITGSSAGLMVAASPVVAMWMTAGQAEITAAQVRVAAAAYETAYGLTVPPVIA 120
DB 61 LTVGSWITGSSAGLMVAASPVVAMWMTAGQAEITAAQVRVAAAAYETAYGLTVPPVIA 120

QY 121 ENRAELMILIAITNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEA 180
DB 121 ENRAELMILIAITNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEA 180

QY 181 PEMSAGGLLEQAAAEEASDTAAANQLMNNVPOALQQLAQPTGTTTSSKLGGLWKTVS 240
DB 181 PEMSAGGLLEQAAAEEASDTAAANQLMNNVPOALQQLAQPTGTTTSSKLGGLWKTVS 240

QY 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSL 300
DB 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSL 300

QY 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAABERGPG 360
DB 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAABERGPG 360

QY 361 QMLGGLPVGMQARAGGSLGVLVRPVPVMPHSPAAAGDIAPPALSDQREDFPALPLD 420
DB 361 QMLGGLPVGMQARAGGSLGVLVRPVPVMPHSPAAAGDIAPPALSDQREDFPALPLD 420

QY 421 PSAMVAQVGPQVNVINTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480
DB 421 PSAMVAQVGPQVNVINTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480

QY 481 QTYGVVDVVDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVPG 540
DB 481 QTYGVVDVVDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVPG 540

QY 541 RVVALGQTVQASDSLTCABEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
DB 541 RVVALGQTVQASDSLTCABEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596

RESULT 2

US-09-886-349A-20
; Sequence 20, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/866,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TBH9-Ra35 (designated MT559F)
US-09-886-349A-20

Query Match

Best Local Similarity 100.0%; Score 3007; DB 12; Length 596;

; Pred. No. 5e-202;

Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HMEHHHHMVDGALPPEINSMRYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWG 60
DB 1 HMEHHHHMVDGALPPEINSMRYAGPGSASLVAAAQMWDSVASDLFSAASAFQSVVWG 60

QY 61 LTVGSWITGSSAGLMVAASPVVAMWMTAGQAEITAAQVRVAAAAYETAYGLTVPPVIA 120
DB 61 LTVGSWITGSSAGLMVAASPVVAMWMTAGQAEITAAQVRVAAAAYETAYGLTVPPVIA 120

QY 121 ENRAELMILIAITNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEA 180
DB 121 ENRAELMILIAITNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEA 180

QY 181 PEMSAGGLLEQAAAEEASDTAAANQLMNNVPOALQQLAQPTGTTTSSKLGGLWKTVS 240
DB 181 PEMSAGGLLEQAAAEEASDTAAANQLMNNVPOALQQLAQPTGTTTSSKLGGLWKTVS 240

QY 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSL 300
DB 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSL 300

QY 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAABERGPG 360
DB 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPOQAWAANAQVTPAARALPLTSLTSAABERGPG 360

QY 361 QMLGGLPVGMQARAGGSLGVLVRPVPVMPHSPAAAGDIAPPALSDQREDFPALPLD 420
DB 361 QMLGGLPVGMQARAGGSLGVLVRPVPVMPHSPAAAGDIAPPALSDQREDFPALPLD 420

QY 421 PSAMVAQVGPQVNVINTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480
DB 421 PSAMVAQVGPQVNVINTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480

QY 481 QTYGVVDVVDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVPG 540
DB 481 QTYGVVDVVDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMGNSGGQGTTPRAVPG 540

QY 541 RVVALGQTVQASDSLTCABEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
DB 541 RVVALGQTVQASDSLTCABEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596

RESULT 3

US-10-359-460-26
; Sequence 26, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1

```
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-10-359-460-26

Query Match      100.0%; Score 3007; DB 14; Length 596;
Best Local Similarity 100.0%; Pred. No. 5e-202;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HMEHHHHMVDYFGALPPEINSARMYAGPGASLVAAAQWDSVASDLFSAASAFQSVVWG 60
DB 1 HMEHHHHMVDYFGALPPEINSARMYAGPGASLVAAAQWDSVASDLFSAASAFQSVVWG 60

QY 61 LTVGSWIGSSAGLVAAASPYVAMSVTAQAELETAQVRAAAAYETAYGLTVPPPVIA 120
DB 61 LTVGSWIGSSAGLVAAASPYVAMSVTAQAELETAQVRAAAAYETAYGLTVPPPVIA 120

QY 121 ENRAELMILIAINLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATILLPPEEA 180
DB 121 ENRAELMILIAINLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATILLPPEEA 180

QY 181 PEMSAGLLEQAAAEEASDTAAANOLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVS 240
DB 181 PEMSAGLLEQAAAEEASDTAAANOLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVS 240

QY 241 PHRSPISNNVMYMANHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSL 300
DB 241 PHRSPISNNVMYMANHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSL 300

QY 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGGP 360
DB 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGGP 360

QY 361 QMLGGLPVQMGARAGGGLSGVLRPVPPVYVPHSPAAGDIAPPALSQDRFADFPALPLD 420
DB 361 QMLGGLPVQMGARAGGGLSGVLRPVPPVYVPHSPAAGDIAPPALSQDRFADFPALPLD 420

QY 421 PSAMVAQVGPQVNVNINIKLYNNNAVAGTGIVDPNGVVLTNHVIAGATDINAFSVGSG 480
DB 421 PSAMVAQVGPQVNVNINIKLYNNNAVAGTGIVDPNGVVLTNHVIAGATDINAFSVGSG 480

QY 481 QTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVAVMGNSSGGQGTTPRAVPG 540
DB 481 QTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVAVMGNSSGGQGTTPRAVPG 540

QY 541 RVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
DB 541 RVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
```

RESULT 4

```
US-10-098-732A-20
; Sequence 20, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; NUMBER OF SEQ ID NOS: 837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-10-098-732A-20

Query Match      100.0%; Score 3007; DB 14; Length 596;
Best Local Similarity 100.0%; Pred. No. 5e-202;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HMEHHHHMVDYFGALPPEINSARMYAGPGASLVAAAQWDSVASDLFSAASAFQSVVWG 60
DB 1 HMEHHHHMVDYFGALPPEINSARMYAGPGASLVAAAQWDSVASDLFSAASAFQSVVWG 60

QY 61 LTVGSWIGSSAGLVAAASPYVAMSVTAQAELETAQVRAAAAYETAYGLTVPPPVIA 120
DB 61 LTVGSWIGSSAGLVAAASPYVAMSVTAQAELETAQVRAAAAYETAYGLTVPPPVIA 120

QY 121 ENRAELMILIAINLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATILLPPEEA 180
DB 121 ENRAELMILIAINLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATILLPPEEA 180

QY 181 PEMSAGLLEQAAAEEASDTAAANOLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVS 240
DB 181 PEMSAGLLEQAAAEEASDTAAANOLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVS 240

QY 241 PHRSPISNNVMYMANHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSL 300
DB 241 PHRSPISNNVMYMANHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSL 300

QY 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGGP 360
DB 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERGGP 360

QY 361 QMLGGLPVQMGARAGGGLSGVLRPVPPVYVPHSPAAGDIAPPALSQDRFADFPALPLD 420
DB 361 QMLGGLPVQMGARAGGGLSGVLRPVPPVYVPHSPAAGDIAPPALSQDRFADFPALPLD 420

QY 421 PSAMVAQVGPQVNVNINIKLYNNNAVAGTGIVDPNGVVLTNHVIAGATDINAFSVGSG 480
DB 421 PSAMVAQVGPQVNVNINIKLYNNNAVAGTGIVDPNGVVLTNHVIAGATDINAFSVGSG 480

QY 481 QTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVAVMGNSSGGQGTTPRAVPG 540
DB 481 QTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVAVMGNSSGGQGTTPRAVPG 540

QY 541 RVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
DB 541 RVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
```

RESULT 5

```
US-10-369-983-21
; Sequence 21, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

```

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72F
US-10-369-983-21

```

Query Match	98.0%;	Score 2946;	DB 15;	Length 729;
Best Local Similarity	100.0%;	Pred. No. 1.2e-197;		
Matches 588;	Conservative 0;	Mismatches 10;	Indels 0;	Gaps 0;
Qy	9	MVDFGALPPEINSARMYAGPGSASLVAAAQWDSVADLFSAAAFOSVVMGLTVGSMIG	68	
Db	142	MVDFGALPPEINSARMYAGPGSASLVAAAQWDSVADLFSAAAFOSVVMGLTVGSMIG	201	
Qy	69	SSAGLMVAASPYVANKSVTAGQBELTAAQVRVAAAAAYETAYGLTVPPVIAENRAELMI	128	
Db	202	SSAGLMVAASPYVANKSVTAGQBELTAAQVRVAAAAAYETAYGLTVPPVIAENRAELMI	261	
Qy	129	LIATNLGQNTPAIVANAEYBGENWADAAAFGYAAATATATATLFFPEAPEMTSAGG	189	
Db	262	LIATNLGQNTPAIVANAEYBGENWADAAAFGYAAATATATATLFFPEAPEMTSAGG	321	
Qy	189	LLSQAAAVEASDPAANQNLNNVPOALQQLAQPTQGTTPSSKLGGLWKTVSPHRSPLN	248	
Db	322	LLSQAAAVEASDPAANQNLNNVPOALQQLAQPTQGTTPSSKLGGLWKTVSPHRSPLN	381	
Qy	249	MVSMANNHMTNSGVSMNTLTSMLKGFAPAAAAQVQTAQNGVRAMSSLSGSSGSSG	308	
Db	382	MVSMANNHMTNSGVSMNTLTSMLKGFAPAAAAQVQTAQNGVRAMSSLSGSSGSSG	441	
Qy	309	LGGVVAANLGRASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERPGQWMLGGLPV	368	
Db	442	LGGVVAANLGRASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERPGQWMLGGLPV	501	
Qy	369	QMGARAGGSLGVLRVPPRPYVPHSPAGDIAPPALSQDRFADFPALPLDPSAMVAQV	428	
Db	502	QMGARAGGSLGVLRVPPRPYVPHSPAGDIAPPALSQDRFADFPALPLDPSAMVAQV	561	
Qy	429	GPQVNVINTKLGYNNAVAGGTGIVDPNGVVLNNHVIAGATDINAFSVSGSGGTGYGVDV	488	
Db	562	GPQVNVINTKLGYNNAVAGGTGIVDPNGVVLNNHVIAGATDINAFSVSGSGGTGYGVDV	621	
Qy	489	GYDRTQDVAVLQIRGAGGLPSAALGGGVAVGEPVVMKNSGGGCTPRVPGRVVVALQGT	548	
Db	622	GYDRTQDVAVLQIRGAGGLPSAALGGGVAVGEPVVMKNSGGGCTPRVPGRVVVALQGT	681	
Qy	549	VQASDSLTAAGETLNGLIQFDAAIQPGDSGGPVYVNGLVGVGMNTAAS	596	
Db	682	VQASDSLTAAGETLNGLIQFDAAIQPGDSGGPVYVNGLVGVGMNTAAS	729	

```

RESULT 6
US-10-369-983-15
; Sequence 15, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-0090811US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB81F (MTB72E-DPV)

```

Query Match 98.0%; Score 2946; DB 15; Length 825;
Best Local Similarity 100.0%; Pred. No. 1.4e-197; Indels 0; Gaps 0;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MYDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGWSIG 68
DB 142 MYDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGWSIG 201

QY 69 SSAGLMVAASPYVAMSVTACQAEELTAQVRAAAAYETAYGLTPPPVIAENRAELMI 128
DB 202 SSAGLMVAASPYVAMSVTACQAEELTAQVRAAAAYETAYGLTPPPVIAENRAELMI 261

QY 129 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 188
DB 262 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 321

QY 189 LLEQAAAVEREASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 248
DB 322 LLEQAAAVEREASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 381

QY 249 MYSVANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAACQNGVRAMSSLGSLGSSG 308
DB 382 MYSVANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAACQNGVRAMSSLGSLGSSG 441

QY 309 LGGVVAANLGRAASVGSLSVPOQAAAAQAVTPAARALPLTSLTSAABERGQMLGGLPV 368
DB 442 LGGVVAANLGRAASVGSLSVPOQAAAAQAVTPAARALPLTSLTSAABERGQMLGGLPV 501

QY 369 GQMGARAGGGLSGVLRVPPRPVMPHSPAAAGDIAPALSDQRFADFPALPLDPSAMVAQV 428
DB 502 GQMGARAGGGLSGVLRVPPRPVMPHSPAAAGDIAPALSDQRFADFPALPLDPSAMVAQV 561

QY 429 GPQVNVINTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVVDV 488
DB 562 GPQVNVINTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVVDV 621

QY 489 GYDRTQDVAVLQLRGAGGLPSAAGGVAVGEPVAVMGNSGGQGTTPRAVGRVVALGQT 548
DB 622 GYDRTQDVAVLQLRGAGGLPSAAGGVAVGEPVAVMGNSGGQGTTPRAVGRVVALGQT 681

QY 549 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWMTAAS 596
DB 682 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWMTAAS 729

RESULT 8
US-10-369-983-13
; Sequence 13, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 875
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MT89F (MTB72F-Erd14)
US-10-369-983-13

Query Match 98.0%; Score 2946; DB 15; Length 875;
Best Local Similarity 100.0%; Pred. No. 1.5e-197;

Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MYDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGWSIG 68
DB 142 MYDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGWSIG 201

QY 69 SSAGLMVAASPYVAMSVTACQAEELTAQVRAAAAYETAYGLTPPPVIAENRAELMI 128
DB 202 SSAGLMVAASPYVAMSVTACQAEELTAQVRAAAAYETAYGLTPPPVIAENRAELMI 261

QY 129 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 188
DB 262 LIATNLLGONTPTAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 321

QY 189 LLEQAAAVEREASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 248
DB 322 LLEQAAAVEREASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 381

QY 249 MYSVANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAACQNGVRAMSSLGSLGSSG 308
DB 382 MYSVANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAACQNGVRAMSSLGSLGSSG 441

QY 309 LGGVVAANLGRAASVGSLSVPOQAAAAQAVTPAARALPLTSLTSAABERGQMLGGLPV 368
DB 442 LGGVVAANLGRAASVGSLSVPOQAAAAQAVTPAARALPLTSLTSAABERGQMLGGLPV 501

QY 369 GQMGARAGGGLSGVLRVPPRPVMPHSPAAAGDIAPALSDQRFADFPALPLDPSAMVAQV 428
DB 502 GQMGARAGGGLSGVLRVPPRPVMPHSPAAAGDIAPALSDQRFADFPALPLDPSAMVAQV 561

QY 429 GPQVNVINTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVVDV 488
DB 562 GPQVNVINTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVVDV 621

QY 489 GYDRTQDVAVLQLRGAGGLPSAAGGVAVGEPVAVMGNSGGQGTTPRAVGRVVALGQT 548
DB 622 GYDRTQDVAVLQLRGAGGLPSAAGGVAVGEPVAVMGNSGGQGTTPRAVGRVVALGQT 681

QY 549 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWMTAAS 596
DB 682 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWMTAAS 729

RESULT 9
US-10-098-732A-65
; Sequence 65, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: US 60/275,837
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB72F-MAPS
; OTHER INFORMATION: (r95f) fusion construct, TB MTB72F (Ra12-TbH9-Ra35)
; OTHER INFORMATION: linked to Leishmania thiol-specific antioxidant
; OTHER INFORMATION: (TSA or MAPS)
US-10-098-732A-65

Query Match 98.0%; Score 2946; DB 14; Length 930;

Best Local Similarity 100.0%; Pred. No. 1.6e-197;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	9	MVDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIG	68
Db	142	MVDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIG	201
QY	69	SSAGLMTVAASPYVAMSVTAGQAEITAAQVRAAAAYETAYGLTVPPPIAENRAELMI	128
Db	202	SSAGLMTVAASPYVAMSVTAGQAEITAAQVRAAAAYETAYGLTVPPPIAENRAELMI	261
QY	129	LIATNLGQNTPAIYNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPEMTSAGG	188
Db	262	LIATNLGQNTPAIYNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPEMTSAGG	321
QY	189	LEQAAAVEASDTAAANQLMNVPOALQOLAQTOGTTTSSKLGGLWKTIVSPHRSPISN	248
Db	322	LEQAAAVEASDTAAANQLMNVPOALQOLAQTOGTTTSSKLGGLWKTIVSPHRSPISN	381
QY	249	MVSMANNHMTNSGYSMVTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG	308
Db	382	MVSMANNHMTNSGYSMVTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG	441
QY	309	LGCGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV	368
Db	442	LGCGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV	501
QY	369	QMGARAGGGLSVLRVPRPYVPHSPAAAGDIAPPALSQDRPADFPALPLDPSAMVAQV	428
Db	502	QMGARAGGGLSVLRVPRPYVPHSPAAAGDIAPPALSQDRPADFPALPLDPSAMVAQV	561
QY	429	GPQVNVNTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDV	488
Db	562	GPQVNVNTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDV	621
QY	489	GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAMGNSGGQGTTPRAVGRVVALGQT	548
Db	622	GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAMGNSGGQGTTPRAVGRVVALGQT	681
QY	549	VQASDSLTGAETLNGLIQFDDAIQPGDSGGPVVNGLGQVGMNTAAS	596
Db	682	VQASDSLTGAETLNGLIQFDDAIQPGDSGGPVVNGLGQVGMNTAAS	729

RESULT 10

US-10-369-983-12
; Sequence 12, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: R95F (MTB72F-MAPS)
US-10-369-983-12

Query Match 98.0%; Score 2946; DB 15; Length 930;
Best Local Similarity 100.0%; Pred. No. 1.6e-197;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	9	MVDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIG	68
Db	142	MVDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIG	201
QY	69	SSAGLMTVAASPYVAMSVTAGQAEITAAQVRAAAAYETAYGLTVPPPIAENRAELMI	128
Db	202	SSAGLMTVAASPYVAMSVTAGQAEITAAQVRAAAAYETAYGLTVPPPIAENRAELMI	261
QY	129	LIATNLGQNTPAIYNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPEMTSAGG	188
Db	262	LIATNLGQNTPAIYNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPEMTSAGG	321
QY	189	LEQAAAVEASDTAAANQLMNVPOALQOLAQTOGTTTSSKLGGLWKTIVSPHRSPISN	248
Db	322	LEQAAAVEASDTAAANQLMNVPOALQOLAQTOGTTTSSKLGGLWKTIVSPHRSPISN	381
QY	249	MVSMANNHMTNSGYSMVTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG	308
Db	382	MVSMANNHMTNSGYSMVTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG	441
QY	309	LGCGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV	368
Db	442	LGCGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV	501
QY	369	QMGARAGGGLSVLRVPRPYVPHSPAAAGDIAPPALSQDRPADFPALPLDPSAMVAQV	428
Db	502	QMGARAGGGLSVLRVPRPYVPHSPAAAGDIAPPALSQDRPADFPALPLDPSAMVAQV	561
QY	429	GPQVNVNTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDV	488
Db	562	GPQVNVNTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDV	621
QY	489	GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAMGNSGGQGTTPRAVGRVVALGQT	548
Db	622	GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAMGNSGGQGTTPRAVGRVVALGQT	681
QY	549	VQASDSLTGAETLNGLIQFDDAIQPGDSGGPVVNGLGQVGMNTAAS	596
Db	682	VQASDSLTGAETLNGLIQFDDAIQPGDSGGPVVNGLGQVGMNTAAS	729

RESULT 11

US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85B)
US-10-369-983-18

Query Match 98.0%; Score 2946; DB 15; Length 1016;
Best Local Similarity 100.0%; Pred. No. 1.8e-197;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	69	SSAGIMVAAASPYYAMNSVTTAGQAEILTAQVRVAAAAVETAYGLTVPPVPVIAENRAELMI	128
Db	202	SSAGIMVAAASPYYAMNSVTTAGQAEILTAQVRVAAAAVETAYGLTVPPVPVIAENRAELMI	261
QY	129	LIATNLLGQNTPAITAVNEAEYGEWMQAQDAAMFGYAAATATATATLLPPEEAPEMTSAGG	188
Db	262	LIATNLLGQNTPAITAVNEAEYGEWMQAQDAAMFGYAAATATATATLLPPEEAPEMTSAGG	321
QY	199	LLEQAAAVEEASDTPAAANQLMNVPQALQOQAOTGCTPSSKLGGLWKVTUSPHRSPISN	248
Db	322	LLEQAAAVEEASDTPAAANQLMNVPQALQOQAOTGCTPSSKLGGLWKVTUSPHRSPISN	381
QY	249	MVSMANNHMSWTNSGVSWNTLTLSMLKGFPAPAAAQAVQTAQNGVRAMSSILGSSLGSSG	308
Db	382	MVSMANNHMSWTNSGVSWNTLTLSMLKGFPAPAAAQAVQTAQNGVRAMSSILGSSLGSSG	441
QY	309	LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAAEALPLTSLTSAAREGPPQMLGGLPV	368
Db	442	LGGGVAANLGRAASVGSLSVPQAWAANQAVTPAAEALPLTSLTSAAREGPPQMLGGLPV	501
QY	369	QOMGARAGGSLUGVLVPPRPYPNPHSPAAAGDTAPPALSDQRFADPPALPLDPSAMVAQV	428
Db	502	QOMGARAGGSLUGVLVPPRPYPNPHSPAAAGDTAPPALSDQRFADPPALPLDPSAMVAQV	561
QY	429	GPQVNNITKLGYNNAVGAGTGIVIPDNGVVLTNNHVIAQTDINAFSVGSSQTTGVDDV	488
Db	562	GPQVNNITKLGYNNAVGAGTGIVIPDNGVVLTNNHVIAQTDINAFSVGSSQTTGVDDV	621
QY	489	GYDRTQDVAVTLQIRGAGGLPSAATGGGVAVGEPPVAMGNSGGQGTTPRAVPPGRVVALGQT	548
Db	622	GYDRTQDVAVTLQIRGAGGLPSAATGGGVAVGEPPVAMGNSGGQGTTPRAVPPGRVVALGQT	681
QY	549	VQASDLSLTGABETLNGLIQFDAALQPGDSGGPVVNGLGQVVGKNTAAS	596
Db	682	VQASDLSLTGABETLNGLIQFDAALQPGDSGGPVVNGLGQVVGKNTAAS	729

```

RESULT 13
US-10-369-903-16
; Sequence 16, Application US/10369993
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB114F (MTB72F-mTCC#2)
US-10-369-903-16

```

Query Match 98.0%; Score 2946; DB 15; Length 1154;
Best Local Similarity 100.0%; Pred. No. 2.2e-197;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 9 MYDFGALPEPEINSRMYAGPGCSASLVAAAQWMDVSADILFSAASQSVVWGLTVGWSWG 68
142 MYDFGALPEPEINSRMYAGPGCSASLVAAAQWMDVSADILFSAASQSVVWGLTVGWSWG 201
Db
QY 69 SSAGLWVAAAQSPYYVAMWSVTAGQRELTAAQVRVAAAAAYETAYGLTVPPFVPIAENRAELMI 128

202 SSAGLWVAASPYVAMVSVTAGQAELETAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 261
129 LIATNLLGQNTPAIAYNEABYEGEMWQDAADAAAFGYAAATATATATALLPPEEAPMTSAGG 188
262 LIATNLLGQNTPAIAYNEABYEGEMWQDAADAAAFGYAAATATATATALLPPEEAPMTSAGG 321
189 LLEQAAAEEASDTAAANQNMNNVPQALQLOAQTOCTTPSSKLGGLWKTVPSPHRSPI SN 248
322 LLEQAAAEEASDTAAANQNMNNVPQALQLOAQTOCTTPSSKLGGLWKTVPSPHRSPI SN 381
249 MVSMAHHMSTNGSVMTNTLSMMLKGFAPAAAAQAVQTAQAQNGVRAVMSLSSSLGSSG 308
382 MVSMAHHMSTNGSVMTNTLSMMLKGFAPAAAAQAVQTAQAQNGVRAVMSLSSSLGSSG 441
309 LGGVVAANLGRAASVGLSVPOAWAANAQAVTPAARALPLTSLTSAAERPGQMLGGLPV 368
442 LGGVVAANLGRAASVGLSVPOAWAANAQAVTPAARALPLTSLTSAAERPGQMLGGLPV 501
369 GQMGARAGGGLSVLRVPPRYVMPHSPAAAGDIAPPALSDQDRFADFPALPLDPSAMVAQV 428
502 GQMGARAGGGLSVLRVPPRYVMPHSPAAAGDIAPPALSDQDRFADFPALPLDPSAMVAQV 561
429 GPQVNNINTKLGYNNAVAGTGIIVDPNGVLTNNHVIAGATDINAFSVSGGQTYGVDVV 488
562 GPQVNNINTKLGYNNAVAGTGIIVDPNGVLTNNHVIAGATDINAFSVSGGQTYGVDVV 621
489 GYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGNSGGQGTTPRAVPGRVVWALGOT 548
622 GYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGNSGGQGTTPRAVPGRVVWALGOT 681
549 VQASDSLTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWMTAAS 596
682 VQASDSLTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWMTAAS 729

RESULT 14

US-09-886-349A-18
; Sequence 18, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Steven
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72FMutSA
; OTHER INFORMATION: (Ra12-TbHp-Ra35MutSA)
US-09-886-349A-18

Query Match 97.9%; Score 2943; DB 12; Length 729;
Best Local Similarity 99.8%; Pred. No. 1.9e-197;
Matches 587; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
9 MVDGALPPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
142 MVDGALPPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSMIG 201
69 SSAGLWVAASPYVAMVSVTAGQAELETAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 128

202 SSAGLWVAASPYVAMVSVTAGQAELETAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 261
129 LIATNLLGQNTPAIAYNEABYEGEMWQDAADAAAFGYAAATATATATALLPPEEAPMTSAGG 188
262 LIATNLLGQNTPAIAYNEABYEGEMWQDAADAAAFGYAAATATATATALLPPEEAPMTSAGG 321
189 LLEQAAAEEASDTAAANQNMNNVPQALQLOAQTOCTTPSSKLGGLWKTVPSPHRSPI SN 248
322 LLEQAAAEEASDTAAANQNMNNVPQALQLOAQTOCTTPSSKLGGLWKTVPSPHRSPI SN 381
249 MVSMAHHMSTNGSVMTNTLSMMLKGFAPAAAAQAVQTAQAQNGVRAVMSLSSSLGSSG 308
382 MVSMAHHMSTNGSVMTNTLSMMLKGFAPAAAAQAVQTAQAQNGVRAVMSLSSSLGSSG 441
309 LGGVVAANLGRAASVGLSVPOAWAANAQAVTPAARALPLTSLTSAAERPGQMLGGLPV 368
442 LGGVVAANLGRAASVGLSVPOAWAANAQAVTPAARALPLTSLTSAAERPGQMLGGLPV 501
369 GQMGARAGGGLSVLRVPPRYVMPHSPAAAGDIAPPALSDQDRFADFPALPLDPSAMVAQV 428
502 GQMGARAGGGLSVLRVPPRYVMPHSPAAAGDIAPPALSDQDRFADFPALPLDPSAMVAQV 561
429 GPQVNNINTKLGYNNAVAGTGIIVDPNGVLTNNHVIAGATDINAFSVSGGQTYGVDVV 488
562 GPQVNNINTKLGYNNAVAGTGIIVDPNGVLTNNHVIAGATDINAFSVSGGQTYGVDVV 621
489 GYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGNSGGQGTTPRAVPGRVVWALGOT 548
622 GYDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGNSGGQGTTPRAVPGRVVWALGOT 681
549 VQASDSLTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWMTAAS 596
682 VQASDSLTGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWMTAAS 729

RESULT 15

US-10-098-732A-18
; Sequence 18, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72FMutSA
; OTHER INFORMATION: (Ra12-TbH9-Ra35MutSA)
US-10-098-732A-18

Query Match 97.9%; Score 2943; DB 14; Length 729;
Best Local Similarity 99.8%; Pred. No. 1.9e-197;
Matches 587; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
9 MVDGALPPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
142 MVDGALPPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVVWGLTVGSMIG 201
69 SSAGLWVAASPYVAMVSVTAGQAELETAQVRVAAAAYETAYGLTVPPVPIAENRAELMI 128

Db	202	SSAGLNVAAASPYPVWMSVTAGQBELTAAQVRVAAAAAYETAYGLTVPPVIAENRAELMI	261
QY	129	LIATNLLQONTPTAIAVNEAEYGEWMAQDAAAMFGYAAATATATATATLPPFEAPEMTSAGG	188
Db	262	LIATNLLQONTPTAIAVNEAEYGEWMAQDAAAMFGYAAATATATATLPPFEAPEMTSAGG	321
QY	189	LLEQAAVEEASDTPAAANCLMNNVPQALQOLAQPTQGTTPSSKLGGLWKTVPHRSPISN	248
Db	322	LLEQAAVEEASDTPAAANCLMNNVPQALQOLAQPTQGTTPSSKLGGLWKTVPHRSPISN	381
QY	249	MVSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAAQAVCTAAQNGVRAMSSLSGSSG	308
Db	382	MVSMANNHMSMTNSGVSMNTLSSMLKGFAPAAAAQAVCTAAQNGVRAMSSLSGSSG	441
QY	309	LGCGVAANLGRAASVGSLSVPQAAHAAANOAVTPAARALPLTSLTSAABERGQMLGGLPV	369
Db	442	LGCGVAANLGRAASVGSLSVPQAAHAAANOAVTPAARALPLTSLTSAABERGQMLGGLPV	501
QY	369	GOMGARAGGGLSGVLRVPPRPYVWPHSPAAGDIAPPALSQDRFADFPALPLDPSAMVAQV	428
Db	502	GOMGARAGGGLSGVLRVPPRPYVWPHSPAAGDIAPPALSQDRFADFPALPLDPSAMVAQV	561
QY	429	GPQVNVINTKLYNNNAVAGGTGIVIDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVV	488
Db	562	GPQVNVINTKLYNNNAVAGGTGIVIDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVV	621
QY	489	GYDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMGNSGGGTTPRAVPGRVVALGQT	548
Db	622	GYDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMGNSGGGTTPRAVPGRVVALGQT	681
QY	549	VOASDSLTAABETLNLGIQFDAAIQFDSGGPVPVNGLQGVVGMNTAAS	596
Db	682	VOASDSLTAABETLNLGIQFDAAIQFDSGGPVPVNGLQGVVGMNTAAS	729

Search completed: June 22, 2004, 18:08:01
Job time : 56.2012 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:07:54 ; Search time 3.42692 Seconds
(without alignments)
1446.225 Million cell updates/sec

Title: US-09-886-349A-22
Perfect score: 481
Sequence: 1 VAMSLTVGAGVASADPVDVAV.....PGAAQYIGLVESVAGSCNNY 96

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
- 4: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
- 5: /cgn2_6/ptodata/2/iaa/PTCUS_COMB.pep:*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	481	100.0	96	3	US-08-818-112-102
2	481	100.0	96	4	US-08-818-111-97
3	481	100.0	96	4	US-09-056-556-102
4	481	100.0	96	4	US-09-072-596-97
5	481	100.0	96	4	US-09-072-967-102
6	481	100.0	110	4	US-09-050-739-151
7	419.5	87.2	331	4	US-09-223-040-4
8	419.5	87.2	331	4	US-09-287-843-4
9	419.5	87.2	368	4	US-09-287-849-24
10	419.5	87.2	433	4	US-09-287-849-14
11	419.5	87.2	856	4	US-09-287-849-12
12	419	87.1	299	4	US-09-287-849-19
13	419	87.1	600	4	US-09-287-849-22
14	419	87.1	710	4	US-09-287-849-16
15	153	32.2	30	4	US-10-005-920-5
16	132	27.4	26	4	US-10-005-920-4
17	125	26.0	24	4	US-10-005-920-1
18	121	25.2	24	4	US-10-005-920-2
19	107	22.2	22	4	US-10-005-920-3
20	104	21.6	20	3	US-08-818-112-120
21	104	21.6	20	4	US-08-818-111-115
22	104	21.6	20	4	US-09-056-556-120
23	104	21.6	20	4	US-09-072-596-115
24	104	21.6	20	4	US-09-072-967-120
25	93	19.3	20	3	US-08-818-113-54
26	93	19.3	20	4	US-08-818-111-54
27	93	19.3	20	4	US-09-056-556-54

28	93	19.3	20	4	US-09-072-596-54	Sequence 54, Appl
29	93	19.3	20	4	US-09-072-967-54	Sequence 54, Appl
30	74	15.4	1454	4	US-09-328-352-5793	Sequence 5793, Ap
31	73	15.2	626	2	US-08-596-300A-7	Sequence 7, Appl
32	73	15.2	626	2	US-08-596-300A-14	Sequence 14, Appl
33	72	15.0	314	1	US-08-040-753-2	Sequence 2, Appl
34	72	15.0	386	4	US-09-331-924-9	Sequence 9, Appl
35	72	15.0	386	4	US-10-147-951B-9	Sequence 9, Appl
36	72	15.0	393	4	US-09-331-924-8	Sequence 8, Appl
37	72	15.0	393	4	US-10-147-951B-8	Sequence 8, Appl
38	70.5	14.7	467	4	US-09-684-855-167	Sequence 167, App
39	70.5	14.7	541	4	US-09-252-991A-20524	Sequence 20524, A
40	70.5	14.7	1042	3	US-08-928-361B-11	Sequence 11, Appl
41	70.5	14.7	1042	4	US-09-588-995A-11	Sequence 11, Appl
42	70.5	14.7	1043	3	US-08-928-361B-30	Sequence 30, Appl
43	70.5	14.7	1721	3	US-08-700-851-5	Sequence 5, Appl
44	70.5	14.7	1721	3	US-08-928-361B-6	Sequence 6, Appl
45	70.5	14.7	1721	4	US-09-588-995A-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-818-112-102
; Sequence 102, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-102

Query Match 100.0%; Score 481; DB 3; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.2e-50;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	VAMSLTVGAGVASADPVD	AVINTTCNYGQVVAALNATD	PGAAAQFNAPVAGSYLRNF	LA	60
Db	1	VAMSLTVGAGVASADPVD	AVINTTCNYGQVVAALNATD	PGAAAQFNAPVAGSYLRNF	LA	60
Qy	61	APPQRAAMAAQIQAPV	GAAQYIGLVESVAGS	CNNY	96	
Db	61	APPQRAAMAAQIQAPV	GAAQYIGLVESVAGS	CNNY	96	

RESULT 2
 US-08-818-111-97
 ; Sequence 97, Application US/08818111
 ; Patent No. 6338852
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonia
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 ; TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 148
 ; CORRESPONDENCE ADDRESS:
 ;

RESULT 3
US-09-056-556-1.02
; Sequence 102, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillion, Devin C
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
 TREATMENT OF HIV INFECTION
 NUMBER OF SEQUENCES: 241
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: United States

```

RESULT 4
US-09-072-596-97
; Sequence 97, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 97:
SEQUENCE CHARACTERISTICS:
LENGTH: 96 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-596-97

Query Match 100.0%; Score 481; DB 4; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.2e-50;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQFNASPVQAQSYLRNFLA 60
DB 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQFNASPVQAQSYLRNFLA 60

QY 61 APPQRAAMAAQLOAVPGAQYIGLVESVAGSCNNY 96
DB 61 APPQRAAMAAQLOAVPGAQYIGLVESVAGSCNNY 96

RESULT 5
US-09-072-967-102
Sequence 102, Application US/09072967
Patent No. 6592877
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESS: SEED AND BERRY LLP
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 96 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-967-102

Query Match 100.0%; Score 481; DB 4; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.2e-50;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQFNASPVQAQSYLRNFLA 60
DB 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQFNASPVQAQSYLRNFLA 60

QY 61 APPQRAAMAAQLOAVPGAQYIGLVESVAGSCNNY 96
DB 61 APPQRAAMAAQLOAVPGAQYIGLVESVAGSCNNY 96

RESULT 6
US-09-050-739-151
Sequence 151, Application US/09050739
Patent No. 6641814
GENERAL INFORMATION:
APPLICANT: ANDERSEN, Peter
APPLICANT: NIELSEN, Rikke
APPLICANT: OETTINGER, Thomas
APPLICANT: RASMUSSEN, Peter Birk
APPLICANT: ROSENKRANDS, Ida
APPLICANT: WELDSINGH, Karin
APPLICANT: FLORIO, Walter
TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
TITLE OF INVENTION: DERIVED FROM M. TUBERCULOSIS
FILE REFERENCE: 670001-2002.1
CURRENT APPLICATION NUMBER: US/09/050,739
CURRENT FILING DATE: 1998-03-30
EARLIER APPLICATION NUMBER: 0376/97
EARLIER FILING DATE: 1997-04-02
EARLIER APPLICATION NUMBER: 1277/97
EARLIER FILING DATE: 1997-11-10
EARLIER APPLICATION NUMBER: 60/044,624
EARLIER FILING DATE: 1997-04-18
EARLIER APPLICATION NUMBER: 60/070,488
EARLIER FILING DATE: 1998-01-05
NUMBER OF SEQ ID NOS: 173
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 151
LENGTH: 110
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-050-739-151

Query Match 100.0%; Score 481; DB 4; Length 110;
Best Local Similarity 100.0%; Pred. No. 2.6e-50;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQFNASPVQAQSYLRNFLA 60
DB 15 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQFNASPVQAQSYLRNFLA 74

QY 61 APPQRAAMAAQLOAVPGAQYIGLVESVAGSCNNY 96
DB 75 APPQRAAMAAQLOAVPGAQYIGLVESVAGSCNNY 110

RESULT 7
US-09-223-040-4
Sequence 4, Application US/09223040
Patent No. 6544522

```

; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-223-040-4

Query Match      87.2%; Score 419.5; DB 4; Length 331;
Best Local Similarity 80.4%; Pred. No. 2.8e-42;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY 5 LTVGAGVASA-----DPVDVAVNTTCNYGQVVAALNATDPGAAAOFNASP 49
|||
Db 129 LTVSVAVSEGKTEKHQIRSTNKLDPVDAVNTTCNYGQVVAALNATDPGAAAOFNASP 188
|||
QY 50 VAQSYLRFNLAAPPORAAQAQVPGAAQYIGLVESVAGSCNNY 96
|||
Db 189 VAQSYLRFNLAAPPORAAQAQVPGAAQYIGLVESVAGSCNNY 235
|||

RESULT 8
US-09-287-849-4
; Sequence 4, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-4

Query Match      87.2%; Score 419.5; DB 4; Length 331;
Best Local Similarity 80.4%; Pred. No. 2.8e-42;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY 5 LTVGAGVASA-----DPVDVAVNTTCNYGQVVAALNATDPGAAAOFNASP 49
|||
Db 129 LTVSVAVSEGKTEKHQIRSTNKLDPVDAVNTTCNYGQVVAALNATDPGAAAOFNASP 188
|||
QY 50 VAQSYLRFNLAAPPORAAQAQVPGAAQYIGLVESVAGSCNNY 96
|||
Db 189 VAQSYLRFNLAAPPORAAQAQVPGAAQYIGLVESVAGSCNNY 235
|||

RESULT 9
US-09-287-849-4
; Sequence 4, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-4

Query Match      87.2%; Score 419.5; DB 4; Length 331;
Best Local Similarity 80.4%; Pred. No. 2.8e-42;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY 5 LTVGAGVASA-----DPVDVAVNTTCNYGQVVAALNATDPGAAAOFNASP 49
|||
Db 129 LTVSVAVSEGKTEKHQIRSTNKLDPVDAVNTTCNYGQVVAALNATDPGAAAOFNASP 188
|||
QY 50 VAQSYLRFNLAAPPORAAQAQVPGAAQYIGLVESVAGSCNNY 96
|||
Db 189 VAQSYLRFNLAAPPORAAQAQVPGAAQYIGLVESVAGSCNNY 235
|||

RESULT 10
US-09-287-849-14
; Sequence 14, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849

```

QY	5	LTVGAGVASA-----DPDVAIVNTTCNYGQVVAALNATDPGAAAFQFNASP	189
Db	130	LTVSVAVSEGKPTKEHIQIRSTNKLDPDVAIVNTTCNYGQVVAALNATDPGAAAFQFNASP	189
QY	50	VAQSYLRNFLLAAPPPORAAVAAQLQAVPGAAQYIGLVESVAGSCNNY	96
Db	190	VAQSYLRNFLLAAPPPORAAVAAQLQAVPGAAQYIGLVESVAGSCNNY	236
RESULT 12			
US-09-287-849-19			
; Sequence 19, Application US/09287849			
; Patent No. 6627198			
; GENERAL INFORMATION:			
; APPLICANT: Reed, Steven G.			
; APPLICANT: Skeiky, Yasir A.W.			
; APPLICANT: Dillon, Davin C.			
; APPLICANT: Alderson, Mark			
; APPLICANT: Campos-Neto, Antonio			
; APPLICANT: Corixa Corporation			
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens			
; FILE REFERENCE: 014058-009020US			
; CURRENT APPLICATION NUMBER: US/09/287,849			
; CURRENT FILING DATE: 1999-04-07			
; PRIOR APPLICATION NUMBER: US 08/818,112			
; PRIOR FILING DATE: 1997-03-13			
; PRIOR APPLICATION NUMBER: US 08/942,578			
; PRIOR FILING DATE: 1997-10-01			
; PRIOR APPLICATION NUMBER: US 09/025,197			
; PRIOR FILING DATE: 1998-02-18			
; PRIOR APPLICATION NUMBER: US 09/056,556			
; PRIOR FILING DATE: 1998-04-07			
; PRIOR APPLICATION NUMBER: US 09/223,040			
; PRIOR FILING DATE: 1998-12-30			
; NUMBER OF SEQ ID NOS: 46			
; SOFTWARE: Patent in Ver. 2.1			
; SEQ ID NO 19			
; LENGTH: 299			
; TYPE: PRT			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion			
US-09-287-849-19			
Query Match 87.1%; Score 419; DB 4; Length 299;			
Best Local Similarity 100.0%; Pred. No. 2,8e-42;			
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	15	DPDVAIVNTTCNYGQVVAALNATDPGAAAFQFNASPVAAQSYLRNFLLAAPPPORAAVAAQLQ	74
Db	9	DPDVAIVNTTCNYGQVVAALNATDPGAAAFQFNASPVAAQSYLRNFLLAAPPPORAAVAAQLQ	68
QY	75	AVPGAAQYIGLVESVAGSCNNY	96
Db	69	AVPGAAQYIGLVESVAGSCNNY	90
RESULT 13			
US-09-287-849-22			
; Sequence 22, Application US/09287849			
; Patent No. 6627198			
; GENERAL INFORMATION:			
; APPLICANT: Reed, Steven G.			
; APPLICANT: Skeiky, Yasir A.W.			
; APPLICANT: Dillon, Davin C.			
; APPLICANT: Alderson, Mark			
; APPLICANT: Campos-Neto, Antonio			
; APPLICANT: Corixa Corporation			
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens			
; FILE REFERENCE: 014058-009020US			

```

; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

```

```

Query Match      87.1%; Score 419; DB 4; Length 600;
Best Local Similarity 100.0%; Pred. No. 7.1e-42;
Matches      82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      15  DPVDVAVNTTCNYGQVVAALNATDPCGAAQFNASPVAQSYLRNFLAAPPQRAAQAQIQ 74
DB      402  DPVDVAVNTTCNYGQVVAALNATDPCGAAQFNASPVAQSYLRNFLAAPPQRAAQAQIQ 461

```

```

QY      75  AVPGAQYIGLVESVAGSCNNY 96
DB      462  AVPGAQYIGLVESVAGSCNNY 483

```

```

RESULT 14
US-09-287-849-16
; Sequence 16, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 710
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-16

```

```

Query Match      87.1%; Score 419; DB 4; Length 710;
Best Local Similarity 100.0%; Pred. No. 9e-42;

```

```

Matches      82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      15  DPVDVAVNTTCNYGQVVAALNATDPCGAAQFNASPVAQSYLRNFLAAPPQRAAQAQIQ 74
DB      9  DPVDVAVNTTCNYGQVVAALNATDPCGAAQFNASPVAQSYLRNFLAAPPQRAAQAQIQ 68

QY      75  AVPGAQYIGLVESVAGSCNNY 96
DB      69  AVPGAQYIGLVESVAGSCNNY 90

```

```

RESULT 15
US-10-005-920-5
; Sequence 5, Application US/10005920
; Patent No. 6664096
; GENERAL INFORMATION:
; APPLICANT: Zhang, Ying
; TITLE OF INVENTION: Methods for Improved Diagnosis and Treatment of Mycobacterial Infection
; FILE REFERENCE: 03940034aa
; CURRENT APPLICATION NUMBER: US/10/005,920
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: US 60/251,785
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: US 60/294,602
; PRIOR FILING DATE: 2001-06-01
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-005-920-5

```

```

Query Match      32.2%; Score 155; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 6.3e-12;
Matches      30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      15  DPVDVAVNTTCNYGQVVAALNATDPCGAAQ 44
DB      1  DPVDVAVNTTCNYGQVVAALNATDPCGAAQ 30

```

```

Search completed: June 22, 2004, 17:27:37
JOB time : 4.42692 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 17:23:10 ; Search time 8.89146 Seconds
(without alignments)
3048.105 Million cell updates/sec

Title: US-09-886-349A-22

Perfect score: 481
Sequence: 1 VAMSLTVGAGVADPVDVAV.....PGAAQYIGLVESVAGSCNNY 96

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PTC_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PTCUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09D_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	481	100.0	96	12	US-09-886-349A-22
2	481	100.0	96	14	US-10-193-002-97
3	481	100.0	96	14	US-10-084-843-102
4	481	100.0	96	14	US-10-098-732A-22
5	481	100.0	110	9	US-09-791-171-151
6	481	100.0	110	10	US-09-804-980-151
7	481	100.0	110	12	US-09-943-443-151
8	424	88.1	92	14	US-10-098-732A-63
9	423	87.9	813	15	US-10-369-983-15
10	419.5	87.2	331	9	US-09-287-849-4
11	419.5	87.2	331	14	US-10-359-460-4
12	419.5	87.2	331	15	US-10-359-459-4
13	419.5	87.2	368	9	US-09-287-849-24
14	419.5	87.2	368	14	US-10-359-460-24
15	419.5	87.2	433	9	US-09-287-849-14

16	419.5	87.2	433	14	US-10-359-460-14	Sequence 14, Appl
17	419.5	87.2	856	9	US-09-287-849-12	Sequence 12, Appl
18	419.5	87.2	856	14	US-10-359-460-12	Sequence 12, Appl
19	419	87.1	82	12	US-09-872-186-7	Sequence 7, Appl
20	419	87.1	87	14	US-10-098-732A-61	Sequence 61, Appl
21	419	87.1	282	14	US-10-098-732A-77	Sequence 77, Appl
22	419	87.1	283	14	US-10-098-732A-57	Sequence 57, Appl
23	419	87.1	290	14	US-10-098-732A-51	Sequence 51, Appl
24	419	87.1	299	9	US-09-287-849-19	Sequence 19, Appl
25	419	87.1	299	12	US-09-886-349A-47	Sequence 47, Appl
26	419	87.1	299	14	US-10-359-460-19	Sequence 19, Appl
27	419	87.1	299	14	US-10-098-732A-47	Sequence 47, Appl
28	419	87.1	600	9	US-09-287-849-22	Sequence 22, Appl
29	419	87.1	600	14	US-10-359-460-22	Sequence 22, Appl
30	419	87.1	710	9	US-09-287-849-16	Sequence 16, Appl
31	419	87.1	710	12	US-09-886-349A-49	Sequence 49, Appl
32	419	87.1	710	14	US-10-359-460-16	Sequence 16, Appl
33	419	87.1	710	14	US-10-098-732A-49	Sequence 49, Appl
34	155	32.2	30	13	US-10-005-920-5	Sequence 5, Appl
35	132	27.4	26	13	US-10-005-920-4	Sequence 4, Appl
36	125	26.0	24	13	US-10-005-920-1	Sequence 1, Appl
37	121	25.2	24	13	US-10-005-920-2	Sequence 2, Appl
38	107	22.2	22	13	US-10-005-920-3	Sequence 3, Appl
39	104	21.6	20	14	US-10-193-002-115	Sequence 115, App
40	104	21.6	20	14	US-10-084-843-120	Sequence 120, App
41	93	19.3	20	14	US-10-193-002-54	Sequence 54, Appl
42	93	19.3	20	14	US-10-084-843-54	Sequence 54, Appl
43	75	15.6	296	14	US-10-128-714-3095	Sequence 3095, Ap
44	75	15.6	296	14	US-10-128-714-8095	Sequence 8095, Ap
45	75	15.6	324	12	US-10-425-114-47182	Sequence 47182, A

ALIGNMENTS

RESULT 1
US-09-886-349A-22
; Sequence 22, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB8.4 (DPV)
US-09-886-349A-22

Query Match 100.0%; Score 481; DB 12; Length 96;
Best Local Similarity 100.0%; Pred. No. 3.8e-45;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	VAMSLTVGAGVADPVDVAVINTCNVGVVVAALNATDCAAAQAFNAPVAQSVLRFLA	60
Db	1	VAMSLTVGAGVADPVDVAVINTCNVGVVVAALNATDCAAAQAFNAPVAQSVLRFLA	60
Qy	61	APPQRAAMAQAQAVEGAAQYIGLVESVAGSCNNY	96
Db	61	APPQRAAMAQAQAVEGAAQYIGLVESVAGSCNNY	96

RESULT 2
 US-10-193-002-97
 ; Sequence 97, Application US/10193002
 ; Publication No. US20030135026A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; Skeiky, Yasir A.W.
 ; Dillon, Davin C.
 ; Campos-Neto, Antonia
 ; Houghton, Raymond
 ; Vedwick, Thomas S.
 ; Twardzik, Daniel R.
 ; Lodes, Michael J.
 ; Hendrickson, Ronald C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 ; TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 350
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; City: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/193,002
 ; FILING DATE: 10-Jul-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072,596
 ; FILING DATE: 05-MAY-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.417C9
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 97:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 96 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 97:
 US-10-193-002-97

Query Match 100.0%; Score 481; DB 14; Length 96;
 Best Local Similarity 100.0%; Pred. No. 3.8e-45;
 Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAAFNAPSPVAQSYLNFILA 60
 Db 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAAFNAPSPVAQSYLNFILA 60
 QY 61 APPQRAAMAAQLQAVPGAAQYIGLVESVAGSCNNY 96
 Db 61 APPQRAAMAAQLQAVPGAAQYIGLVESVAGSCNNY 96

RESULT 3
 US-10-084-843-102
 ; Sequence 102, Application US/10084843
 ; Publication No. US20030143243A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.

; Skeiky, Yasir A.W.
 ; Dillon, Davin C.
 ; Campos-Neto, Antonio
 ; Houghton, Raymond
 ; Vedwick, Thomas S.
 ; Twardzik, Daniel R.
 ; Lodes, Michael J.
 ; Hendrickson, Ronald C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 ; AND DIAGNOSIS OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 355
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; City: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/084,843
 ; FILING DATE: 25-Feb-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072,967
 ; FILING DATE: 05-MAY-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.411C9
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 102:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 96 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 102:
 US-10-084-843-102

Query Match 100.0%; Score 481; DB 14; Length 96;
 Best Local Similarity 100.0%; Pred. No. 3.8e-45;
 Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAAFNAPSPVAQSYLNFILA 60
 Db 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAAFNAPSPVAQSYLNFILA 60
 QY 61 APPQRAAMAAQLQAVPGAAQYIGLVESVAGSCNNY 96
 Db 61 APPQRAAMAAQLQAVPGAAQYIGLVESVAGSCNNY 96

RESULT 4
 US-10-098-732A-22
 ; Sequence 22, Application US/10098732A
 ; Publication No. US20030175294A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; Brannon, Mark
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
 ; FILE OF INVENTION: Leishmania Antigen
 ; FILE REFERENCE: 014058-012010US
 ; CURRENT APPLICATION NUMBER: US/10/098,732A
 ; CURRENT FILING DATE: 2003-04-29

```

; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB8.4 (DPV)
US-10-098-732A-22

Query Match      100.0%; Score 481; DB 14; Length 96;
Best Local Similarity 100.0%; Pred. No. 3.8e-45;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQAFNASPVAQSYLRNFLA 60
Db 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQAFNASPVAQSYLRNFLA 60

QY 61 APPQRAAQAOLQAVPGAAQYIGLVESVAGSCNNY 96
Db 61 APPQRAAQAOLQAVPGAAQYIGLVESVAGSCNNY 96

RESULT 5
US-09-791-171-151
; Sequence 151, Application US/09791171
; Patent No. US20020094336A1
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: NIELSEN, Rikke
; APPLICANT: OETTINGER, Thomas
; APPLICANT: RASMUSSEN, Peter Birk
; APPLICANT: ROSENKRANDS, Ida
; APPLICANT: WELDINGH, Karin
; APPLICANT: FLORIO, Walter
; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
; FILE REFERENCE: 670001-2002.1
; CURRENT APPLICATION NUMBER: US/09/791,171
; CURRENT FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 09/050,739
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 0376/97
; PRIOR FILING DATE: 1997-04-02
; PRIOR APPLICATION NUMBER: 1277/97
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/044,624
; PRIOR FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: 60/070,488
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 151
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-791-171-151

Query Match      100.0%; Score 481; DB 9; Length 110;
Best Local Similarity 100.0%; Pred. No. 4.5e-45;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQAFNASPVAQSYLRNFLA 60
Db 15 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQAFNASPVAQSYLRNFLA 74

QY 61 APPQRAAQAOLQAVPGAAQYIGLVESVAGSCNNY 96
Db 75 APPQRAAQAOLQAVPGAAQYIGLVESVAGSCNNY 110
```

```

RESULT 6
US-09-804-980-151
; Sequence 151, Application US/09804980
; Publication No. US20030147897A1
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; APPLICANT: Anderson, Peter
; TITLE OF INVENTION: M. Tuberculosis Antigens
; FILE REFERENCE: 670001-2002.4
; CURRENT APPLICATION NUMBER: US/09/804,980
; CURRENT FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 257
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 151
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-804-980-151

Query Match      100.0%; Score 481; DB 10; Length 110;
Best Local Similarity 100.0%; Pred. No. 4.5e-45;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQAFNASPVAQSYLRNFLA 60
Db 15 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQAFNASPVAQSYLRNFLA 74

QY 61 APPQRAAQAOLQAVPGAAQYIGLVESVAGSCNNY 96
Db 75 APPQRAAQAOLQAVPGAAQYIGLVESVAGSCNNY 110

RESULT 7
US-09-943-443-151
; Sequence 151, Application US/09943443
; Publication No. US20030017172A1
; GENERAL INFORMATION:
; APPLICANT: ANDERSEN, Peter
; APPLICANT: NIELSEN, Rikke
; APPLICANT: OETTINGER, Thomas
; APPLICANT: RASMUSSEN, Peter Birk
; APPLICANT: ROSENKRANDS, Ida
; APPLICANT: WELDINGH, Karin
; APPLICANT: FLORIO, Walter
; TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
; FILE REFERENCE: 670001-2002.1
; CURRENT APPLICATION NUMBER: US/09/943,443
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/050,739
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 1277/97
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/044,624
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-04-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/070,488
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 151
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-943-443-151

Query Match      100.0%; Score 481; DB 12; Length 110;
Best Local Similarity 100.0%; Pred. No. 4.5e-45;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQAFNASPVAQSYLRNFLA 60
Db 15 VAMSLTVGAGVASADPVDVAVINTTCNYGQVVAALNATDPGAAQAFNASPVAQSYLRNFLA 74
```



```

; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 4
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-10-359-460-4

Query Match      87.2%; Score 419.5; DB 14; Length 331;
Best Local Similarity 80.4%; Pred. No. 9.5e-38;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY      5 LTVGAGVASA-----DPVDVINTTCNYGQVVAALNATDPGAAAFNASP 49
|||
Db      129 LTVSVAVSEKPTKHIQIRSTNKLDPVDVINTTCNYGQVVAALNATDPGAAAFNASP 188
|||

QY      50 VAQSYLRNFLAAPPQRAAMAALQAVPGAAYIGLVESVAGSCNNY 96
|||
Db      189 VAQSYLRNFLAAPPQRAAMAALQAVPGAAYIGLVESVAGSCNNY 235
|||

RESULT 12
US-10-359-459-4
; Sequence 4, Application US/10359459
; Publication No. US20040013677A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/10/359,459
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 4
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-10-359-459-4

Query Match      87.2%; Score 419.5; DB 15; Length 331;
Best Local Similarity 80.4%; Pred. No. 9.5e-38;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY      5 LTVGAGVASA-----DPVDVINTTCNYGQVVAALNATDPGAAAFNASP 49
|||

```

```

Db      129 LTVSVAVSEKPTKHIQIRSTNKLDPVDVINTTCNYGQVVAALNATDPGAAAFNASP 188
|||
QY      50 VAQSYLRNFLAAPPQRAAMAALQAVPGAAYIGLVESVAGSCNNY 96
|||
Db      189 VAQSYLRNFLAAPPQRAAMAALQAVPGAAYIGLVESVAGSCNNY 235
|||

RESULT 13
US-09-287-849-24
; Sequence 24, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 24
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-24

Query Match      87.2%; Score 419.5; DB 9; Length 368;
Best Local Similarity 80.4%; Pred. No. 1.1e-37;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY      5 LTVGAGVASA-----DPVDVINTTCNYGQVVAALNATDPGAAAFNASP 49
|||
Db      130 LTVSVAVSEKPTKHIQIRSTNKLDPVDVINTTCNYGQVVAALNATDPGAAAFNASP 189
|||

QY      50 VAQSYLRNFLAAPPQRAAMAALQAVPGAAYIGLVESVAGSCNNY 96
|||
Db      190 VAQSYLRNFLAAPPQRAAMAALQAVPGAAYIGLVESVAGSCNNY 236
|||

RESULT 14
US-10-359-460-24
; Sequence 24, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849

```

;; PRIOR FILING DATE: 1999-04-07
;; PRIOR APPLICATION NUMBER: US 08/818,112
;; PRIOR FILING DATE: 1997-03-13
;; PRIOR APPLICATION NUMBER: US 08/942,578
;; PRIOR FILING DATE: 1997-10-01
;; PRIOR APPLICATION NUMBER: US 09/025,197
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 09/056,556
;; PRIOR FILING DATE: 1998-04-07
;; PRIOR APPLICATION NUMBER: US 09/223,040
;; PRIOR FILING DATE: 1998-12-30
;; NUMBER OF SEQ ID NOS: 46
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 24
;; LENGTH: 368
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-10-359-460-24

Query Match 87.2%; Score 419.5; DB 14; Length 368;
Best Local Similarity 80.4%; Pred. No. 1.1e-37;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY 5 LTVGAGVASA-----DPVDVAVNTTCNYGQVVAALNATDPGAAAFNASP 49
|||
DB 130 LTVSVAVSEKPTKHKIQRSTNKLDPVDVAVNTTCNYGQVVAALNATDPGAAAFNASP 189
|||
QY 50 VAQSYLRNFLAAPPORAAQAQVPGAAQYIGLVESVAGSCNNY 96
|||
DB 190 VAQSYLRNFLAAPPORAAQAQVPGAAQYIGLVESVAGSCNNY 236
|||

RESULT 15
US-09-287-849-14
; Sequence 14, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-003020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-14

Query Match 87.2%; Score 419.5; DB 9; Length 433;
Best Local Similarity 80.4%; Pred. No. 1.3e-37;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY 5 LTVGAGVASA-----DPVDVAVNTTCNYGQVVAALNATDPGAAAFNASP 49
|||
DB 130 LTVSVAVSEKPTKHKIQRSTNKLDPVDVAVNTTCNYGQVVAALNATDPGAAAFNASP 189
|||
QY 50 VAQSYLRNFLAAPPORAAQAQVPGAAQYIGLVESVAGSCNNY 96
|||
DB 190 VAQSYLRNFLAAPPORAAQAQVPGAAQYIGLVESVAGSCNNY 236
|||

Search completed: June 22, 2004, 18:08:02
Job time : 9.89146 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:44:25 ; Search time 11.7858 Seconds
(without alignments)
2301.458 Million cell updates/sec

Title: US-09-886-349A-22

Perfect score: 481

Sequence: 1 VAMSLTVGAGVASADPVDV.....PCAAQYIGLVSVAGSCNNY 96

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04: *
1: Geneseqp1980s: *
2: Geneseqp1990s: *
3: Geneseqp2000s: *
4: Geneseqp2001s: *
5: Geneseqp2002s: *
6: Geneseqp2003as: *
7: Geneseqp2003bs: *
8: Geneseqp2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	481	100.0	96	2 AAW32349	AAW32349 Mycobacte
2	481	100.0	96	2 AAW32417	AAW32417 Mycobacte
3	481	100.0	96	2 AAW64333	AAW64333 Mycobacte
4	481	100.0	96	2 AAW81700	AAW81700 M. tuberc
5	481	100.0	96	2 AAY38987	AAY38987 M. tuberc
6	481	100.0	96	2 AAY39130	AAY39130 M. tuberc
7	481	100.0	96	4 AAU01894	AAU01894 M. tuberc
8	481	100.0	96	5 AAE29711	AAE29711 Mycobacte
9	481	100.0	96	5 AAE17575	AAE17575 Mycobacte
10	481	100.0	110	2 AAW72936	AAW72936 Mycobacte
11	481	100.0	110	2 AAY21956	AAY21956 Amino aci
12	424	88.1	92	5 AAE29730	AAE29730 Leishmani
13	423	87.9	813	7 ADA26367	ADA26367 Mycobacte
14	419.5	87.2	330	5 AAU74589	AAU74589 Antigenic
15	419.5	87.2	331	2 AAY32060	AAY32060 Mycobacte
16	419.5	87.2	368	5 AAY32069	AAY32069 Mycobacte
17	419.5	87.2	368	5 AAU74598	AAU74598 Antigenic
18	419.5	87.2	433	2 AAY32065	AAY32065 Mycobacte
19	419.5	87.2	433	5 AAU74594	AAU74594 Antigenic
20	419.5	87.2	856	2 AAY32064	AAY32064 Mycobacte
21	419.5	87.2	859	5 AAU74593	AAU74593 Antigenic
22	419	87.1	82	3 AAY96924	AAY96924 M. tuberc
23	419	87.1	87	5 AAE29729	AAE29729 Leishmani
24	419	87.1	282	5 AAE29736	AAE29736 Mycobacte
25	419	87.1	283	5 AAE29727	AAE29727 Leishmani

26	419	87.1	290	5 AAE29725	AAE29725 Leishmani
27	419	87.1	299	2 AAY32067	AAY32067 Mycobacte
28	419	87.1	299	5 AAE29723	AAE29723 Mycobacte
29	419	87.1	299	5 AAE17587	AAE17587 Mycobacte
30	419	87.1	306	5 AAU74596	AAU74596 Antigenic
31	419	87.1	600	2 AAY32068	AAY32068 Mycobacte
32	419	87.1	600	5 AAU74597	AAU74597 Antigenic
33	419	87.1	710	2 AAY32066	AAY32066 Mycobacte
34	419	87.1	710	5 AAE29724	AAE29724 Mycobacte
35	419	87.1	710	5 AAE17588	AAE17588 Mycobacte
36	419	87.1	723	5 AAU74595	AAU74595 Antigenic
37	326	67.8	110	6 AAE37873	AAE37873 Mycobacte
38	229	47.6	60	4 AAB52437	AAB52437 Mycobacte
39	155	32.2	30	5 ABB79492	ABB79492 Mycobacte
40	132	27.4	26	5 ABB79491	ABB79491 Mycobacte
41	125	26.0	24	5 ABB79488	ABB79488 Mycobacte
42	121	25.2	24	5 ABB79489	ABB79489 Mycobacte
43	107	22.2	22	5 ABB79490	ABB79490 Mycobacte
44	104	21.6	20	2 AAW32325	AAW32325 Mycobacte
45	104	21.6	20	2 AAW32394	AAW32394 Mycobacte

ALIGNMENTS

RESULT 1
AAW32349
ID AAW32349 standard; protein; 96 AA.
XX AC AAW32349;
XX AC AAW32349;
DT 13-JAN-1998 (first entry)
XX XX
DE Mycobacterium tuberculosis antigen DPV.
XX XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX XX
OS Mycobacterium tuberculosis.
XX XX
FN WO9709429-A2.
XX XX
PD 13-MAR-1997.
XX XX
PF 30-AUG-1996; 96WO-US014675.
XX XX
PR 01-SEP-1995; 95US-00523435.
PR 22-SEP-1995; 95US-00532136.
PR 22-MAR-1996; 96US-00620280.
PR 05-JUN-1996; 96US-00658800.
PR 12-JUL-1996; 96US-00680573.
XX XX
PA (CORI-) CORIXA CORP.
XX XX
PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedvick TH, Twardzik DR;
XX XX
DR WPI; 1997-192904/17.
DR N-PSDB; AAT91424.
XX XX
PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
PT useful for diagnosis of M. tuberculosis infection.
XX XX
PS Example 1; Page 146; 190pp; English.
XX XX
CC A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
CC variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a M.tuberculosis antigen,
CC DPV. The immunogenic polypeptide can be used to diagnose M.tuberculosis
CC infection by forming complexes with specific antibodies in the sample.
CC Fragments of DNA encoding the immunogenic polypeptide can be used as
CC diagnostic primers or probes and agents that bind to the antigen,

CC especially monoclonal antibodies or equivalent polyclonal antibodies, are
 CC also used for diagnosis

SQ Sequence 96 AA;

Query Match 100.0%; Score 481; DB 2; Length 96;
 Best Local Similarity 100.0%; Pred. No. 8.6e-47;
 Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAQFNASPVAAQSYLRNFLA 60

DB 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAQFNASPVAAQSYLRNFLA 60

QY 61 APPPQRAAQAQIQAVPGAQYIGLVESVAGSCNNY 96

DB 61 APPPQRAAQAQIQAVPGAQYIGLVESVAGSCNNY 96

RESULT 2

AAW32417
 ID AAW32417 standard; protein; 96 AA.

XX AC AAW32417;

XX 06-JAN-1998 (first entry)

XX Mycobacterium tuberculosis antigen DPV.

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

KW skin testing; M.tuberculosis.

XX Mycobacterium tuberculosis.

OS OS

XX WO9709428-A2.

XX 13-MAR-1997.

XX 30-AUG-1996; 96WO-US014674.

XX 01-SEP-1995; 95US-00523436.

XX 22-SEP-1995; 95US-00533634.

XX 22-MAR-1996; 96US-00620874.

XX 05-JUN-1996; 96US-00659683.

XX 12-JUL-1996; 96US-00680574.

XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky VA, Dillon DC, Campos-Neto A, Houghton R;

PI Vedvick TH, Twardzik DR;

XX WPI; 1997-192903/17.

XX N-PSDB; AAT91488.

XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also for
 PT diagnosis.

XX Example 1; Page 134; 168pp; English.

XX A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
 CC variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis antigen,
 CC DPV. The immunogenic protein, and fusion proteins containing one or more
 CC of the proteins or one of the proteins plus ESAT-6, are useful in
 CC vaccines, preferably when formulated with a non-specific adjuvant, to
 CC induce an immune response against M.tuberculosis (for treatment or
 CC prevention)

XX SQ Sequence 96 AA;

Query Match 100.0%; Score 481; DB 2; Length 96;
 Best Local Similarity 100.0%; Pred. No. 8.6e-47;

Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAQFNASPVAAQSYLRNFLA 60

DB 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAQFNASPVAAQSYLRNFLA 60

QY 61 APPPQRAAQAQIQAVPGAQYIGLVESVAGSCNNY 96

DB 61 APPPQRAAQAQIQAVPGAQYIGLVESVAGSCNNY 96

RESULT 3

AAW64333
 ID AAW64333 standard; protein; 96 AA.

XX AC AAW64333;

XX 17-OCT-2003 (revised)

DT 09-NOV-1998 (first entry)

XX Mycobacterium tuberculosis antigen DPV.

XX Tuberculosis; infection; diagnosis; antigen; DPV.

XX Mycobacterium tuberculosis; strain H37Rv.

XX WO9816645-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US018214.

XX 11-OCT-1996; 96US-00729622.

XX 13-MAR-1997; 97US-00818111.

XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;

PI Vedvick TS, Twardzik DR, Lodes MJ;

XX WPI; 1998-251292/22.

XX N-PSDB; AAV44392.

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and
 PT diagnosis of tuberculosis.

XX Example 1; Page 129-130; 250pp; English.

XX This polypeptide comprises the Mycobacterium tuberculosis antigen DPV. A
 CC DNA sequence (see AAV44392) coding for DPV was isolated from a M.
 CC tuberculosis genomic library using a probe based on an isolated N-
 CC terminal peptide (see AAW64341). The invention relates to compositions
 CC and methods for diagnosing tuberculosis. It provides polypeptides (see
 CC AAW64291-W64379) comprising an antigenic portion of a soluble M.
 CC tuberculosis antigen, or an immunogenic portion of an M. tuberculosis
 CC antigen, as well as DNA sequences encoding such polypeptides, recombinant
 CC expression vectors and transfected or transfected host cells. Also
 CC claimed are methods and diagnostic kits for detecting M. tuberculosis
 CC infection in a patient using these polypeptides, antibodies or
 CC oligonucleotide probes and primers, for the diagnosis of tuberculosis.
 CC (Updated on 17-OCT-2003 to standardise OS field)

XX SQ Sequence 96 AA;

Query Match 100.0%; Score 481; DB 2; Length 96;
 Best Local Similarity 100.0%; Pred. No. 8.6e-47;

Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAQFNASPVAAQSYLRNFLA 60

DB 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAQFNASPVAAQSYLRNFLA 60

CC are used in methods for eliciting immune response in mammals. They are
CC useful as vaccines to elicit protective immunity against pathogenic
CC microorganisms such as leishmania and Mycobacterium tuberculosis. Fusion
CC polypeptides are used for enhancing the expression of polynucleotides, as
CC in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is M.
CC tuberculosis DPV antigenic protein. DPV is also referred to as MTB8.4
XX
XX
SQ Sequence 96 AA;

Query Match 100.0%; Score 481; DB 5; Length 96;
Best Local Similarity 100.0%; Pred. No. 8.6e-47;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAAFNAPSPVAQSYLRNFLA 60
DB 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAAFNAPSPVAQSYLRNFLA 60
QY 61 APPQRAAQAQLQAVPGAAQYIGLVESVAGSCNNY 96
DB 61 APPQRAAQAQLQAVPGAAQYIGLVESVAGSCNNY 96

RESULT 9
AAE17575
ID AAE17575 standard; protein; 96 AA.
XX
AC AAE17575;
XX
XX
DT 22-APR-2002 (first entry)
XX
DE Mycobacterium species MTB8.4 (DPV) protein.
XX
XX Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; MTB8.4; DPV protein.
KW
XX
OS Mycobacterium sp.
XX
XX WO200198460-A2.
XX
PD 27-DEC-2001.
XX
XX 20-JUN-2001; 2001WO-US019959.
XX
XX 20-JUN-2000; 2000US-00597796.
PR
XX 01-FEB-2001; 2001US-0265737P.
XX
XX (CORI-) CORIXA CORP.
XX
XX
XX Skeiky Y, Reed S, Alderson M;
PI
XX
XX WPI; 2002-147738/19.
DR
XX N-PSDB; AAD28345.
XX

Composition comprising MTB39 antigen and MTB32A antigen from
PT Mycobacterium species, useful for eliciting immune response in a subject.
XX
XX
XX Claim 9; Page 116; 136pp; English.
XX
XX
XX The present invention relates to fusion proteins containing at least two
CC Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected with
CC Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the

CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC MTB8.4 (DPV) protein
XX
XX
SQ Sequence 96 AA;

Query Match 100.0%; Score 481; DB 5; Length 96;
Best Local Similarity 100.0%; Pred. No. 8.6e-47;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAAFNAPSPVAQSYLRNFLA 60
DB 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPGAAAFNAPSPVAQSYLRNFLA 60
QY 61 APPQRAAQAQLQAVPGAAQYIGLVESVAGSCNNY 96
DB 61 APPQRAAQAQLQAVPGAAQYIGLVESVAGSCNNY 96

RESULT 10
AAW72936
ID AAW72936 standard; protein; 110 AA.
XX
AC AAW72936;
XX
XX 21-JAN-1999 (first entry)
XX
DE Mycobacterium tuberculosis antigen CFP98.
XX
XX Mycobacterium tuberculosis; antigen; vaccine; immunological; immunogen;
KW infection.
KW
XX Mycobacterium tuberculosis.
OS
XX WO9844119-A1.
XX
PD 08-OCT-1998.
XX
XX 01-APR-1998; 98WO-DK000132.
XX
XX 02-APR-1997; 97DK-00000376.
PR
XX 18-APR-1997; 97US-0044624P.
PR
XX 10-NOV-1997; 97DK-00001277.
PR
XX 05-JAN-1998; 98US-0070488P.
XX
XX (STAT-) STATENS SERUM INST.
XX
XX Andersen P, Nielsen R, Rosenkrands I, Weldingh K, Rasmussen PB;
PI Oettinger T, Florio W;
XX
XX WPI; 1998-542705/46.
DR
XX N-PSDB; AAV63946.
XX

New isolated mycobacteria polypeptides and nucleic acids - used for
PT developing products for the diagnosis of or vaccination against
PT mycobacterial infections, particularly tuberculosis.
XX
XX Claim 1; Page 225; 163pp; English.
XX
XX The present sequence represents a Mycobacterium tuberculosis protein.
CC Products from the present invention which describes protein fragments
CC and nucleic acid fragments derived from M.tuberculosis, can be used in
CC the detection of and prevention of mycobacterial infections. In
CC particular, the proteins and nucleic acids can be used for the diagnosis
CC of or vaccination against tuberculosis caused by M. tuberculosis, M.
CC africanum or M. bovis
XX
XX Sequence 110 AA;

Query Match 100.0%; Score 481; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. NO. 1e-46;
 Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPCGAAQFNASPVQAQSYLRNFLA 60
 |||||
 DB 15 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPCGAAQFNASPVQAQSYLRNFLA 74
 |||||

QY 61 APPQRAAAQLOQVPGAAQYIGLVESVAGSCNNY 96
 |||||
 DB 75 APPQRAAAQLOQVPGAAQYIGLVESVAGSCNNY 110
 |||||

RESULT 11

AA21956

ID AAY21956 standard; protein; 110 AA.

XX

AC AAY21956;

XX 06-SEP-1999 (first entry)

XX Amino acid sequence of antigen CFP83.

XX Immunogenic; Mycobacterium tuberculosis; immune response; infection;
 KW tuberculosis; fusion polypeptide; T-cell epitope; ESAT-6; MPT59; TB;
 KW pharmaceutical; vaccination; M. africanum; M. bovis; CFP7A; CFP30A;
 KW CFP7B; CFP19; CFP27; CFP30A; RDI-ORF; CFP10A; CFP16; CFP19; CFP23;
 KW CFP25A; CFP30B; CFP7B.

XX Mycobacterium tuberculosis.

XX WO9924577-A1.

XX 20-MAY-1999.

XX 08-OCT-1998; 98WO-DK000438.

XX 10-NOV-1997; 97DK-00001277.

XX 05-JAN-1998; 98US-0070488P.

XX 01-APR-1998; 98WO-DK000132.

XX (STAT-) STATENS SERUM INST.

XX Andersen P, Skjot R;

XX WPI; 1999-347282/29.

XX N-PSDB; AAX81095.

XX New immunogenic fragment of Mycobacterium tuberculosis.

XX Example 3; Page 242; 265pp; English.

XX The invention describes a substantially pure immunogenic polypeptide
 CC fragment (I) from Mycobacterium tuberculosis that is able to evoke a
 CC protective immune response against infections by mycobacteria belonging
 CC to the tuberculosis complex. The invention provides a (1) fusion
 CC polypeptide comprising at least one polypeptide fragment (I) and at least
 CC one fusion partner; (2) a fusion polypeptide fragment comprising a T-cell
 CC epitope from M. tuberculosis protein ESAT-6, or MPT59 and a second
 CC different amino acid sequence from M. tuberculosis, and/or including a
 CC sequence which protects the first amino acid sequence from in vivo
 CC degradation or post-translational processing; (3) a nucleic acid fragment
 CC that encodes the above polypeptides. The polypeptides and nucleic acid
 CC are useful as pharmaceuticals, for diagnosis of and as antigens for
 CC vaccination against TB caused by Mycobacterium tuberculosis, africanum or
 CC bovis. The polypeptides are also useful for diagnosing ongoing or
 CC previous sensitization in an animal with bacteria belonging to the
 CC tuberculosis complex. The invention also describes the use of CFP7A or
 CC CFP30A or a T-cell epitope of for the induction of a strong immune
 CC response in a mammal; use of CFP7B, CFP19 or MPT59-ESAT6 or a T-cell
 CC epitope of for diagnosis of TB in a mammal by performing a DTH type skin
 CC test; use of CFP27, CFP30A, RDI-ORF2, RDI-ORF3, MPT59-ESAT6,
 CC ESAT6-MPT59, CFP10A, CFP16, CFP19, CFP23, CFP25A, CFP30B, CFP7B or a T-

CC cell epitope of for the preparation of an immunological composition; and
 CC for the preparation of a subunit vaccine

XX Sequence 110 AA;

Query Match 100.0%; Score 481; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. NO. 1e-46;

Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPCGAAQFNASPVQAQSYLRNFLA 60
 |||||

DB 15 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDPCGAAQFNASPVQAQSYLRNFLA 74
 |||||

QY 61 APPQRAAAQLOQVPGAAQYIGLVESVAGSCNNY 96
 |||||

DB 75 APPQRAAAQLOQVPGAAQYIGLVESVAGSCNNY 110
 |||||

RESULT 12

AAE29730

ID AAE29730 standard; protein; 92 AA.

XX

AC AAE29730;

XX 29-AUG-2003 (revised)

XX 27-JAN-2003 (first entry)

XX Leishmania sp. MAPS (N10)-Mycobacterium sp. DPV-AC fusion protein.

XX Vaccine; immunity; diagnostic agent; gene therapy; MAPS; DPV;
 KW fusion protein.

XX Mycobacterium sp.

XX Leishmania sp.

XX Chimeric.

XX WO200272792-A2.

XX 19-SEP-2002.

XX 13-MAR-2002; 2002WO-US008223.

XX 13-MAR-2001; 2001US-0275837P.

XX (CORI-) CORIXA CORP.

XX Skeiky Y, Brannon M, Guderian J;

XX WPI; 2002-759844/82.

XX N-PSDB; AAD47109.

XX New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif.
 CC M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 CC against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 CC tuberculosis.

XX Example 4; Page 127-128; 155pp; English.

XX The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC Leishmania species thiol specific antioxidant [TSA; MAPS (N10)]-
 CC Mycobacterium sp. DPV-AC (codon optimised) fusion protein. This sequence
 CC comprises the first ten amino acids of Leishmania species TSA (MAPS) at

CC the N-terminus linked with the TB antigen DPV (aka MTB8.4) which is codon
CC optimised for expression in eukaryotic cells. (Updated on 29-AUG-2003 to
CC standardise OS field)
XX
SQ Sequence 92 AA;

Query Match 88.1%; Score 424; DB 5; Length 92;
Best Local Similarity 90.2%; Pred. No. 2.5e-40;
Matches 83; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 5 LTVGAGVASADPDVAVINTTCNYGQVVAALNATDPCAAAFNAPVQSVLYRNFLAAPP 64
DB 1 MSCGNKINSDDPDVAVINTTCNYGQVVAALNATDPCAAAFNAPVQSVLYRNFLAAPP 60

QY 65 QRAAMAAQLQAVPGAAQYIGLVESVAGSCNNY 96
DB 61 QRAAMAAQLQAVPGAAQYIGLVESVAGSCNNY 92

RESULT 13
ADA26367
ID ADA26367 standard; protein; 813 AA.
AC ADA26367;
XX
XX
DT 20-NOV-2003 (first entry)
DE
DE Mycobacterium MTB72F-DPV (fusion MTB81F) protein.
KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW tuberculosis; tuberculostatic; gene therapy; vaccine.
XX
XX Chimeric.
OS
OS Mycobacterium sp.
XX
XX WO2003070187-A2.
PN
PN
PD
PD 28-AUG-2003.
XX
XX
PF 18-FEB-2003; 2003WO-US004903.
XX
XX 15-FEB-2002; 2002US-0357351P.
PR
PR (CORI-) CORIXA CORP.
PA
PA Skeiky Y, Guderian J, Reed S;
PI
PI
DR
DR WPI; 2003-697554/66.
DR N-PSDB; ADA26360.
XX
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT and MTB39 antigens, with or without the MTB85A antigen, from a
PT Mycobacterium species, useful for diagnosing, preventing and/or treating
PT tuberculosis infection.
XX
XX
PS Claim 85; Fig 15; 112pp; English.

CC The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.

XX
SQ Sequence 813 AA;

Query Match 87.9%; Score 423; DB 7; Length 813;
Best Local Similarity 92.4%; Pred. No. 4.5e-39;
Matches 85; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 7 VGAGVASA--DPVDVAVINTTCNYGQVVAALNATDPCAAAFNAPVQSVLYRNFLAAPP 64

DB 722 VGMNTAASGTRDPDVAVINTTCNYGQVVAALNATDPCAAAFNAPVQSVLYRNFLAAPP 781
QY 65 QRAAMAAQLQAVPGAAQYIGLVESVAGSCNNY 96
DB 782 QRAAMAAQLQAVPGAAQYIGLVESVAGSCNNY 813

RESULT 14
AAU74589
ID AAU74589 standard; protein; 330 AA.
AC AAU74589;
XX
XX
DT 29-AUG-2003 (revised)
DT 08-MAY-2002 (first entry)
XX
XX Antigenic fusion protein Erd14-DPV-MTI.
XX
XX Fusion protein; tuberculosis; Mycobacterium tuberculosis;
KW tuberculostatic; immunogen; vaccine; Erd14-DPV-MTI; Erd14; DPV; MTI.
XX
XX Mycobacterium tuberculosis.
OS
OS Chimeric.
XX
XX US2002009459-A1.
PN
PN
PD 24-JAN-2002.
XX
XX 07-APR-1999; 99US-00287849.
PF
PF 13-MAR-1997; 97US-00818112.
PR 01-OCT-1997; 97US-00942578.
PR 18-FEB-1998; 98US-00025197.
PR 07-APR-1998; 98US-00056556.
PR 30-DEC-1998; 98US-00223040.
XX
XX (REED/) REED S G.
PA (SKEIKY) SKEIKY Y A.
PA (DILL/) DILLON D C.
PA (ALDE/) ALDERSON M.
PA (CAMP/) CAMPOS-NETO A.
XX
XX Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
PI
PI WPI; 2002-171134/22.
DR N-PSDB; ABK14129.
XX
XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for
PT diagnosing, treating or preventing M. tuberculosis infection,
PT particularly as vaccine for treating or preventing tuberculosis.
XX
XX Claim 1; Fig 2; 62pp; English.

CC The invention relates to a purified polypeptide which induces an immune
CC response of Mycobacterium tuberculosis. Polypeptides of the invention are
CC useful for diagnosing, treating or preventing M. tuberculosis infection,
CC particularly tuberculosis infection. In particular, the polypeptides are
CC useful as a vaccine formulation with an adjuvant to afford long-term
CC protection in animals against the development of tuberculosis. The
CC protein coding sequence may be used to encode a protein product for use
CC as an immunogen to induce and/or enhance an immune response to M.
CC tuberculosis. This sequence represents an M. tuberculosis fusion protein
CC of the invention. (Updated on 29-AUG-2003 to standardise OS field)

XX
SQ Sequence 330 AA;

Query Match 87.2%; Score 419.5; DB 5; Length 330;
Best Local Similarity 80.4%; Pred. No. 3.8e-39;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY 5 LTVGAGVASA-----DPVDVAVINTTCNYGQVVAALNATDPCAAAFNAPVQSVLYRNFLAAPP 49

Db	129	LTVSVAVSGKTEKHXH	QIRSTNKLDPDAVIN	TTCNYGQVVAALNATD	PAGAAQNASP	188
QY	50	VAGSYLRNF	LAAPPOR	AAMAQLOAV	FGAAQY	96
Db	189	VAGSYLRNF	LAAPPOR	AAMAQLOAV	FGAAQY	235

RESULT 15

RESULT IS
 AAY32060
 ID AAY32060 standard; protein; 331 AA.

AA
AC
AAY32060:

17-JAN-2000 (first entry)

DE Mycobacterium tuberculosis antigen fusion protein Mtb39A;

AA Tuberculosis; antigen; fusion protein; Mtb39A; ERD14; DPV; MTI;
KW diagnosis; therapy; vaccine; immunogen.

OS *Mycobacterium tuberculosis*.

Key	Location/Qualifiers
1.7	/note="Met/His tag"
8.151	/note="ERD14"
154.1235	/note="DPV"
238.1331	/note="MTII"

XX PN WO9951748-A2.

14-OCT-1999

07-APR-1999; 99WO-US007717.

XX 07-APR-1998: 98US-00056556

FR 07-AFK-1998; 98US-00036336;
PR 30-DEC-1998; 98US-00223040;

PA (CORI-) CORIXA CORP.

PI Skeiky YAW, Alderson M, Campos-Neto A;

XX
DR
WPI: 1999-601610/51.

DR N-PSDB; AAZ20195.

New fusion proteins useful for diagnosis, prevention and treatment of tuberculosis.

PS Claim 1; Fig 2; 83pp; English.

This sequence represents a recombinant *Mycobacterium tuberculosis* tri-antigen fusion protein, termed Mtb39A, composed of the antigens ERD14, DPV and Mri. The fusion protein is expressed in host cells using a vector carrying a polynucleotide (see AAZ20195) comprising the 3 coding sequences for the antigens. The invention provides fusion proteins (see AAY32059-71) containing at least 2 M. tuberculosis antigens. The new fusion proteins and polynucleotides encoding them are useful as vaccines for preventing tuberculosis (claimed), for diagnosis (via in vitro assays or intradermal skin tests for detection of anti-M. tuberculosis antibodies), monitoring of disease progression, and treatment of tuberculosis. They are more effective immunogens than mixtures of the individual protein components.

Sequence 331 AA;

Query Match 87.2%; Score 419.5; DB 2; Length 331;
Best Local Similarity 80.4%; Pred. No. 3.8e-39;
Matches 86; Conservative 1; Mismatches 5; Indels 15; Gaps 1;

QY 5 LTVGAGVASA-----DPVDAVINTTCNYGOVWAALNATDPGAAAOFNASP 49

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:59:04 ; Search time 2.68596 Seconds
(without alignments)
3438.018 Million cell updates/sec

Title: US-09-886-349A-22

Perfect score: 481
Sequence: 1 VAMSLTVGAGVASADPVDV.....PGAAQYIGLVESVAGSCNNY 96

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	481	100.0	110	2 F70875	hypothetical prote
2	73.5	15.3	1477	2 AG3009	polyketide synthet
3	73.5	15.3	1489	2 G98274	hypothetical prote
4	73	15.2	626	2 S36364	glucan 1,4-alpha-g
5	73	15.2	626	2 T49625	glucan 1,4-alpha-g
6	72.5	15.1	197	2 A6430	F26G16.12 protein
7	72	15.0	196	2 A81966	probable periplasm
8	72	15.0	344	2 A70771	probable hydrolase
9	72	15.0	345	1 VABPA7	major capsid prote
10	72	15.0	398	1 VBBPA7	minor capsid prote
11	70.5	14.7	462	2 B31123	probable endoprote
12	70.5	14.7	1832	2 T31113	mucin-like glycopr
13	70	14.6	196	2 D81022	conserved hypothet
14	70	14.6	451	2 D83395	protein F53A3.6 li
15	69.5	14.4	288	2 G84351	thiamin monophosph
16	69	14.3	227	2 G70555	hypothetical prote
17	69	14.3	244	2 D96707	probable zinc fing
18	69	14.3	343	1 D33858	mercury(II) reduct
19	69	14.3	655	2 AD1598	DNA gyrase-like pr
20	69	14.3	655	2 AF1235	DNA gyrase-like pr
21	68.5	14.2	362	2 G87470	SlyP family secret
22	68	14.1	243	2 T29624	spherulin lb prote
23	68	14.1	347	2 T46192	acetylglutamate ki
24	67.5	14.0	326	2 AD3269	malate dehydrogena
25	67.5	14.0	865	2 A47282	calcium-binding pr
26	67.5	14.0	1400	2 T31555	hypothetical prote
27	67	13.9	465	2 G02738	FRFAC-4 - human
28	67	13.9	1607	2 G87259	conserved hypothet
29	67	13.9	1607	2 T02837	long chain fatty a

30 66.5 13.8 314 2 T03554 potential transcri
31 66.5 13.8 353 2 B94393 D-lactate dehydrog
32 66.5 13.8 399 2 S53975 probable membrane
33 66.5 13.8 732 2 A40839 ornithine decarbox
34 66.5 13.8 842 2 S18462 glycoprotein H pre
35 66 13.7 228 2 T36379 probable peptidase
36 66 13.7 351 2 C83089 histidinol-phospha
37 66 13.7 1226 2 T24045 hypothetical prote
38 66 13.7 1226 2 B97243 probable cation tr
39 66 13.7 3013 2 AB0480 probable invasiv y
40 66 13.7 3184 1 WMBEH6 UL36 protein - hum
41 65.5 13.6 186 2 T03570 hypothetical prote
42 65.5 13.6 287 2 G86728 alpha-subuni L-ser
43 65.5 13.6 484 2 A83082 probable porin PA4
44 65.5 13.6 663 2 S55164 hypothetical prote
45 65 13.5 332 2 A75352 hypothetical prote

ALIGNMENTS

RESULT 1

F70875
hypothetical protein Rv1174c - Mycobacterium tuberculosis (strain H37RV)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C/Accession: F70875
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: F70875
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-110 <COL>
A/Cross-references: GB:AL010186; GB:AL123456; NID:g3261493; PIDN:CAA15951.1; PID:g2695955
A/Experimental source: strain H37RV
C/Genetics:
A/Gene: Rv1174c
C/Superfamily: Mycobacterium tuberculosis hypothetical protein Rv1174c

Query Match 100.0%; Score 481; DB 2; Length 110;
Best Local Similarity 100.0%; Pred. No. 28-41;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 VAMSLTVGAGVASADPVDVINTTCNYGVVAALNATDPGAAQFNASPVQAQSLRNFLA.60
Db 15 VAMSLTVGAGVASADPVDVINTTCNYGVVAALNATDPGAAQFNASPVQAQSLRNFLA.74
Qy 61 APPPQRAAQAQVPGAAQYIGLVESVAGSCNNY 96
Db 75 APPPQRAAQAQVPGAAQYIGLVESVAGSCNNY 110

RESULT 2

AG3009
polyketide synthetase Atu3681 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C/Accession: AG3009
R/Wood, D.W.; Serubai, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan; Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, J.; ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193
A/Accession: AG3009
A/Status: preliminary

C;Accession: A86430
R;Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.;
ausen, N.F.; Hughes, B.; Huijzer, L.
Nature 408, 816-820, 2000

A;Authors: Salzbberg, S.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Razzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzbberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A86430

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-197 <STO>
A;Cross-references: GB:AE005172; NID:G6634774; PIDN:AAE19754.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Query Match 15.1%; Score 72.5; DB 2; Length 197;
Best Local Similarity 31.0%; Pred. No. 4.3;
Matches 26; Conservative 8; Mismatches 23; Indels 27; Gaps 5;

QY 26 NYGQVVAALNATDPG-----AAQFNASPVQSVLRNFLAAPPQ-----RAAMAAQLQ 74
DB 41 NYESIVTSLVSDPGTNSMAPGQY---PYDPYYS-IFAPPQPYTGVLHQLMGVQQQ 96

QY 75 AVP-----GAAQYIGLV 86
DB 97 GVPLPSDAVEPFVFNKQYHGIL 120

RESULT 7
E81966
probable periplasmic transport protein NMA0488 [imported] - Neisseria meningitidis (stra
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: E81966

R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagers, K.; Leather, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000

A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: E81966

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-196 <PAR>
A;Cross-references: GB:AL162753; GB:AL157959; NID:G7379120; PIDN:CAB83781.1; PID:G737923
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: NMA0488

Query Match 15.0%; Score 72; DB 2; Length 196;
Best Local Similarity 32.6%; Pred. No. 4.8;
Matches 28; Conservative 11; Mismatches 43; Indels 4; Gaps 2;

QY 10 GVASADPVDVAVTTCNVGVVAALNATDPGAA---AQFNASPVQSVLRNFLAAPPQ 66
DB 18 GMAFAAPADAVNQIRQNAQVLSILKSGDANTARQAEAYAIYPDFQFMTALAVGNPW 77

QY 67 AAMAAQLQAVPAAQVIGLVESVAGS 92
DB 78 TASDAQKQALAKEFQTL-LIRYSQT 102

RESULT 8
A70771
probable hydrolase - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: A70771

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998

A;Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome s
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70771

A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-344 <COL>
A;Cross-references: GB:Z73902; GB:AL123456; NID:G2261576; PIDN:CAA98097.1; PID:G245027; I
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: Rv1333

Query Match 15.0%; Score 72; DB 2; Length 344;
Best Local Similarity 34.1%; Pred. No. 8.7;
Matches 29; Conservative 7; Mismatches 29; Indels 20; Gaps 3;

QY 2 AMSLTVGAGVASAD-----PVDVAVTTCNVGVVAALNATDPGAAQFNASPVQSV 54
DB 155 AGALKGGVGTASATLQSGVTGVLAVN-----AAGNVVDPATGLPMMADLVGEFA 205

QY 55 LRNFLAAPPQRAAMAAQLQAVPGA 79
DB 206 LR-----APPAEQIALAQJSSPLGA 226

major capsid protein 10A - phase T7
C;Species: phage T7
A;Note: host Escherichia coli
C;Date: 13-Jun-1983 #sequence_revision 30-Sep-1990 #text_change 23-Jul-1999
C;Accession: A04344; S42325
R;Dunn, J.J.; Thompson, K.
submitted to the Nucleic Acid Sequence Database, September 1982

A;Reference number: A94615
A;Accession: A04344
A;Molecule type: DNA
A;Residues: 1-345 <DUN>
R;Dunn, J.J.; Studier, F.W.
J. Mol. Biol. 166, 477-535, 1983

A;Title: Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 ge
A;Reference number: S42283; MUID:83241725; PMID:6864790
A;Accession: S42325
A;Molecule type: DNA
A;Residues: 1-345 <DUW>
A;Cross-references: EMBL:V01146; NID:G431187; PIDN:CAA24427.1; PID:G15604
A;Note: the authors did not translate the codon for residue 1
C;Genetics:
A;Gene: 10A

A;Map position: 57.51-60.49
C;Superfamily: phage T7 major capsid protein 10A
C;Keywords: capsid protein

Query Match 15.0%; Score 72; DB 1; Length 345;
Best Local Similarity 31.6%; Pred. No. 8.8;
Matches 31; Conservative 8; Mismatches 35; Indels 24; Gaps 5;

QY 8 GAGVASADPVDVAVTTCNVG-----QVVAALNATDPGAAQFNASPVQSVLRN 57
DB 158 GLGTAT-----VIETQNKAAITDQVALGKEIIAAL--TKARAALTKNVPAAD---RV 206

QY 58 FLAAPPQRAAMAAQLQAVPGAQYIGLVESVAGSCNN 95
DB 207 FYCDPSYSAILAA---LMPNAAAYALIDPEKGSIRN 241

RESULT 10
VBPA7
minor capsid protein 10B - phase T7
C;Species: phage T7

A>Note: host Escherichia coli
C:Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 23-Jul-1999
C:Accession: B04344; S42326
R:Dunn, J.J.; Thompson, K.
submitted to the Nucleic Acid Sequence Database, September 1982
A:Reference number: A94615
A:Accession: B04344
A:Molecule type: DNA
A:Residues: 1-398 <DUN>
R:Dunn, J.J.; Studier, F.W.
J. Mol. Biol. 166, 477-535, 1983
A:Title: Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genes
A:Reference number: S42283; MUID:83241725; PMID:6864790
A:Accession: S42326
A:Status: preliminary; translation not shown
A:Molecule type: DNA
A:Residues: 1-398 <DU2>
A:Cross-references: EMBL:V01146; NID:G431187; PIDN:CAA24428.1; PID:G431193
C:Genetics:
A:Gene: 108
A:Map position: 57.51-60.49
A>Note: translation of the nucleotide sequence involves a -1 frameshift within codon 341
C:Superfamily: phage T7 major capsid protein 10A
C:Keywords: capsid protein

Query Match 15.0%; Score 72; DB 1; Length 398;
Best Local Similarity 31.6%; Pred. No. 10;
Matches 31; Conservative 8; Mismatches 35; Indels 24; Gaps 5;

QY 8 GAGVAGADPVDVAVNTTCNYG-----QVVAALNATDPCGAAQFNASPVASQYLRN 57
DB 158 GLGTAT-----VIETTONKAAITDQVALGKEIIAAL--TKARAALTKNYVPAAD---RV 206

QY 58 FLAAPPQBAWAAQLOAVPGAAQYIGLVESVAGSCNN 95
DB 207 FYCDPSYSALIAA---LMPNAAVAAALIDPEKGSIRN 241

RESULT 11
B83123
Probable endoproteinase Arg-C precursor PA4175 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83123
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 405, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: B82950; MUID:20437337; PMID:10984043
A:Accession: B83123
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-462 <STO>
A:Cross-references: GB:AE004834; GB:AE004091; NID:G9950382; PIDN:AAG07562.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4175

Query Match 14.7%; Score 70.5; DB 2; Length 462;
Best Local Similarity 29.1%; Pred. No. 17;
Matches 23; Conservative 9; Mismatches 34; Indels 13; Gaps 2;

QY 20 VINTTCNYGQVVAALNATD-PCGAAQFNASPVASQYLR-----NFLAAPPPQR 66
DB 12 VLALAGASQALAPGASMGADGVVLQSPASTGHAFANFNAAISAAGIHFAAPPARR 71

QY 67 AAMAAQLOAVPGAAQYIGL 85
DB 72 VARAPLAKPTPLQGV 90

RESULT 12

T31113
mucin-like glycoprotein 900 - Cryptosporidium parvum
C:Species: Cryptosporidium parvum
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31113
R:Barnes, D.A.; Bonnin, A.; Huang, J.X.; Gousset, L.; Wu, J.; Gut, J.; Doyle, P.; Dubremet, M.
Mol. Biochem. Parasitol. 96, 93-110, 1998
A:Title: A novel multi-domain mucin-like glycoprotein of Cryptosporidium parvum mediates adherence to host cells
A:Reference number: Z20989; MUID:99066935; PMID:9851610
A:Accession: T31113
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1832 <BAR>
A:Cross-references: EMBL:AF068065; NID:G4063041; PID:G4063042; PIDN:AAC98153.1

Query Match 14.7%; Score 70.5; DB 2; Length 1832;
Best Local Similarity 28.4%; Pred. No. 74;
Matches 21; Conservative 12; Mismatches 26; Indels 15; Gaps 2;

QY 31 VAALNATDPCGAAQFNASPVASQY-----LRNFLAA-----PPQRAAQAALQQA 75
DB 1690 VSGVFATVGTATPKKGVIPESVAAEAARKYFAANVEGEGEEVPPPESSSNTAIOA 1749

QY 76 VPGAQYIGLVESV 89
DB 1750 AGGASAAVGLVAAV 1763

RESULT 13
D81022
Conserved hypothetical protein NMB1963 [imported] - Neisseria meningitidis (strain MC58)
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: D81022
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizzi, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappelli, R.; Verma, A.; Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: D81022
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-196 <TET>
A:Cross-references: GB:AE002544; GB:AE002098; NID:G7227214; PIDN:AAF42292.1; PID:G7227222;
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1963

Query Match 14.8%; Score 70; DB 2; Length 196;
Best Local Similarity 32.8%; Pred. No. 7.6;
Matches 28; Conservative 10; Mismatches 44; Indels 4; Gaps 2;

QY 10 GVASADPVDVAVNTTCNYGQVVAALNATDPCGAA---AQFNASPVASQYLRNFLAAPPPQR 66
DB 18 GMAFAPADAVSQIRQATQVLSILKNGDANTARQAEAYAIYFFDFQMTALAVGNPWR 77

QY 67 AAMAAQLOAVPGAAQYIGLVESVAGS 92
DB 78 TASDAQKQALAKEFQTL-LIRTSYSGT 102

RESULT 14
D88395
Protein F53A3.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: D88395
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:9851916

Search completed: June 22, 2004, 17:24:58
Job time : 5.68596 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:46:13 ; Search time 1.59768 Seconds
(without alignments)
3128.737 Million cell updates/sec

Title: US-09-886-349A-22

Perfect score: 481

Sequence: 1 VAMSLTVGAGVASADPVDV.....FGARQYIGLVESVAGSCNNY 96

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	75.5	15.7	621	VP40_HSVBC	P54817 bovine herp
2	73.5	15.3	509	TSSP_MOUSE	Q9qxe5 mus musculu
3	73	15.2	626	AMYG_NEUCR	P14804 neurospora
4	72	15.0	344	YJ33_MYCTU	Q10644 mycobacteri
5	72	15.0	345	VCAA_BPT7	P19726 bacterioph
6	72	15.0	398	VCAB_BPT7	P19727 bacterioph
7	69	14.3	460	MEGA_SERMA	P08662 serratia ma
8	68	14.1	183	REGA_RHOSU	O82868 rhodorum
9	67.5	14.0	246	SRIA_RHYPO	P03350 physarum po
10	67.5	14.0	865	CPN_DROME	Q02910 drosophila
11	67	13.9	465	FXD1_HUMAN	Q16676 homo sapien
12	66.5	13.8	389	SCWA_YEAST	Q04951 saccharomyc
13	66.5	13.8	514	TSSP_HUMAN	Q9qxe7 homo sapien
14	66.5	13.8	732	DCOS_ECOLI	P24169 escherichia
15	66.5	13.8	842	VGLH_HSVBC	P27599 bovine herp
16	66	13.7	351	H181_PSEAE	Q9nwx0 pseudomonas
17	66	13.7	1609	CTPI_MYCLE	O53114 mycobacteri
18	66	13.7	3164	TEGU_HSV11	P10220 herpes simp
19	65.5	13.6	663	YJ09_YEAST	P47005 saccharomyc
20	65	13.5	323	MIAA_RALSO	Q8xbw0 ralstonia s
21	65	13.5	618	VE1_HPV63	Q7847 human papil
22	65	13.5	659	AGEA_ECOLI	P37127 escherichia
23	65	13.5	1157	SRA4_HUMAN	Q95104 homo sapien
24	65	13.5	1442	FTCI_CHICK	Q90693 gallus gall
25	65	13.5	2038	F5H_DROME	P13709 drosophila
26	64.5	13.4	248	SRIE_RHYPO	Q05489 physarum po
27	64.5	13.4	358	LIP_PSEGL	P09351 pseudomonas
28	64.5	13.4	474	NOA1_RAT	Q80wa4 rattus norv
29	64.5	13.4	493	NOA1_MOUSE	Q9jkn6 mus musculu
30	64.5	13.4	510	NOA1_HUMAN	P51513 homo sapien
31	64	13.3	203	CUTI_PRRER	Q9y7g8 pyrenopeziz
32	64	13.3	686	EOMD_HUMAN	O95936 homo sapien
33	64	13.3	1434	PTC1_MOUSE	Q61115 mus musculu

ALIGNMENTS

RESULT 1

ID	VP40_HSVBC	STANDARD;	PRT;	621 AA.
AC	P54817; Q89855;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Capsid protein P40 [Contains: Assemblin (Protease) (EC 3.4.21.97); Capsid assembly protein].			
DE	Capsid assembly protein].			
GN	UL26.			
OS	Bovine herpesvirus type 1 (strain Cooper).			
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
OC	Alphaherpesvirinae; Varicelloviruses.			
OX	NCBI_TaxID=10323;			
[1]				
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=96013856; PubMed=7474173;			
RA	Haanes E.J., Thomsen D.R., Martin S., Homa F.L., Lowery D.E.;			
RT	"The bovine herpesvirus 1 maturational proteinase and scaffold proteins can substitute for the homologous herpes simplex virus type 1 proteins in the formation of hybrid type B capsids.";			
RL	J. Virol. 69:7375-7379(1995).			
CC	-!- FUNCTION: THE CAPSID ASSEMBLY PROTEIN IS COMPONENT OF THE CAPSID CORE INVOLVED IN PROCESSING AND PACKAGING OF PROGENY DNA.			
CC	-!- FUNCTION: ASSEMBLIN IS A PROTEASE WHICH CAN PROTEOLYTICALLY CLEAVE ITSELF AND THE CAPSID ASSEMBLY PROTEIN AT THE C-TERMINUS.			
CC	-!- CATALYTIC ACTIVITY: Cleaves -Ala- -Ser- and -Ala- -Ala- bonds in the scaffold protein.			
CC	-!- ALTERNATIVE PRODUCTS:			
CC	Event=Alternative initiation;			
CC	Comment=2 isoforms, UL26 (shown here) and UL26.5, are produced by alternative initiation;			
CC	-!- SIMILARITY: Belongs to peptidase family S21.			
CC	-!- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP40.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sb-sib.ch).			
CC	-----			
DR	EMBL; U31809; AAA91999.1; -			
DR	EMBL; U31809; AAA92000.1; -			
DR	EMBL; Z78205; CAB01599.1; -			
DR	EMBL; AJ004801; CAA06107.1; -			
DR	HSSP; P16753; 1CMW.			
DR	MEROPS; S21.001; -			
DR	InterPro; IPR001847; Peptidase S21.			
DR	Pfam; PF00716; Peptidase S21; 1.			
DR	PRINTS; PR00236; HSVCAPSIDP40.			
DR	Capsid assembly; Hydrolase; Serine protease; Coat protein;			
KW	Alternative initiation.			
FT	CHAIN 1 621			
FT	CHAIN 314 621			
FT	CAPSID PROTEIN P40, ISOFORM UL26.			
FT	CAPSID PROTEIN P40, ISOFORM UL26.5.			

Q9hb15 homo sapien
Q9ers1 rattus norv
Q00650 brassica na
Q10900 mycobacteri
P28955 equine herp
Q9a4c4 caulobacter
O83525 t ispd/ispf
P17047 mus musculu
P29836 bovine herp
Q9u3v5 drosophila
P34288 caenorhabdi
P09631 mus musculu

FT INIT MET 314 314 FOR ISOFORM UL26.5.
 FT CHAIN 1 254 ASSEMBLIN (PROTEASE).
 FT CHAIN 255 621 CARSID ASSEMBLY PROTEIN.
 FT SITE 254 255 CLEAVAGE (BY THE PROTEASE). (PROBABLE).
 FT ACT SITE 68 68 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 136 136 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 155 155 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SQ SEQUENCE 621 AA; 63711 MW; AEF6D7C9BD569003 CRC64;
 Query Match 15.7%; Score 75.5; DB 1; Length 621;
 Best Local Similarity 33.0%; Pred. No. 2.6;
 Matches 35; Conservative 6; Mismatches 30; Indels 35; Gaps 5;
 QY 2 AMSLTGAGVASADPV---DAVINTTCYQGVV-----AALNATDPGAAAFNAPSPVAQ 52
 Db 313 AMSAAHQAGASAHPLFAGDYVVVPTAQYNQLVVSQARGAMTAAPPAPYFLPAAAA- 371
 QY 53 SYLRNLFAPPPQ-----RAMAA-----QLQAVPGA 79
 Db 372 -----AAAPPMPGMYGAAGAPHPHYGFPFPPGLESQTMALAGA 411

RESULT 2

TSRP_MOUSE STANDARD; PRT; 509 AA.
 AC Q9QXES;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thymus-specific serine protease precursor (EC 3.4.-.-).
 GN PRS16 OR TSSP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Embryonic thymus;
 RX MEDLINE=20093985; PubMed=10630289;
 RA Carrier A., Nguyen C., Victorero G., Granjeaud S., Rocha D.,
 RA Bernard K., Mizak A., Ferrier P., Malissen M., Naquet P.,
 RA Malissen B., Jordan B.R.
 RT "Differential gene expression in CD3epsilon- and RAG1-deficient
 RT thymuses: definition of a set of genes potentially involved in
 RT thymocyte maturation."
 RL Immunogenetics 50:255-270(1999).
 CC -/- FUNCTION: Protease that may play a role in T-cell development.
 CC -/- SUBCELLULAR LOCATION: Vesicular, either lysosomal or endosomal.
 CC -/- TISSUE SPECIFICITY: Expressed predominantly in cortical thymic
 CC epithelial cells, with highest expression around vessels and the
 CC thymic capsule.
 CC -/- DEVELOPMENTAL STAGE: Expressed in developing thymus at 14 to 18
 CC dpc, with maximal expression at 16 dpc.
 CC -/- SIMILARITY: Belongs to peptidase family S28.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AJ131775; CAB66137.1;
 DR MEROPS; S28.003;
 DR MGD; MGI:1859181; Prs16.
 DR InterPro; IPR008758; Peptidase_S28.
 DR InterPro; IPR000379; Ser_estr.
 DR Pfam; PF05577; Peptidase_S28; 1.
 DR KEGG; K05577; Serine protease; Signal.
 FT SIGNAL 1 22 POTENTIAL,
 FT CHAIN 23 509 THYMUS-SPECIFIC SERINE PROTEASE.
 FT ACT_SITE 184 184 CHARGE RELAY SYSTEM (POTENTIAL).

FT ACT SITE 446 446 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 471 471 CHARGE RELAY SYSTEM (POTENTIAL).
 FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 509 AA; 54523 MW; CD8425F3F71BE43C CRC64;
 Query Match 15.3%; Score 73.5; DB 1; Length 509;
 Best Local Similarity 29.2%; Pred. No. 3.4;
 Matches 31; Conservative 13; Mismatches 37; Indels 25; Gaps 5;
 QY 7 VGAGVASADPVADVINTTCYQGVV-ALNATDPG-----AAAQ-----FNAPSV 50
 Db 201 VFAAVASSAPLSAVVDFSA-YNQVVARSLTQVAIGGSLECLAAASTAFTVERLLRAGA 259
 QY 51 ACSYLNRFLAAPPQRAAMAAQLQAVPQAAQYIGLVESVAGSCNNY 96
 Db 260 AQAVIREELGA-----CGSLDLEDAELGALQALVGGTVQY 297
 RESULT 3
 AMVG_NEUCR STANDARD; PRT; 626 AA.
 AC P14804; Q9P5U5;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Glucoamylase precursor (EC 3.2.1.3) (Glucon 1,4-alpha-glucosidase)
 DE (1,4-alpha-D-glucan glucohydrolase).
 GN GLA-1 OR B5022.70.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=74-OR23-1A / FGSC 987;
 RX MEDLINE=94037144; PubMed=8221928;
 RA Stone P.J., Makoff A.J., Parish J.H., Radford A.;
 RA "Cloning and sequence analysis of the glucoamylase gene of Neurospora
 RA crassa."
 RL Curr. Genet. 24:205-211(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=74-OR23-1A / FGSC 987;
 RX MEDLINE=22542210; PubMed=12655011;
 RA Mannheim G., Montrose C., Haase D., Mewes H.-W., Aign V.,
 RA Hoheisel J.D., Fartmann B., Nyakatura G., Kempken F., Maier J.,
 RA Schulte U.;
 RT "What's in the genome of a filamentous fungus? Analysis of the
 RT Neurospora genome sequence."
 RL Nucleic Acids Res. 31:1944-1954(2003).
 RN [3]
 RP SEQUENCE OF 36-65.
 RC STRAIN=74-OR23-1A / FGSC 987;
 RA Koh-Luar S.I., Parish J.H., Bleasby A.J., Pappin D.J.C., Ainley K.,
 RA Johansen P.E., Radford A.;
 RT "Exported proteins of Neurospora crassa: 1-glucoamylase."
 RL Enzyme Microb. Technol. 11:692-695(1989).
 CC -/- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -/- SIMILARITY: Belongs to family 15 of glycosyl hydrolases.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X67291; CAA47707.1;
 DR

```

DR EMBL; AL355932; CAB91426.1; -.
DR HSSP; P04064; 1GAI.
DR InterPro; IPR002044; CBD 4.
DR InterPro; IPR008291; Glu-a-glcsd_SBD.
DR InterPro; IPR00165; Glyco_hydro_15.
DR InterPro; IPR008928; Glyco_trans_6hp.
DR Pfam; PF00586; CSM_20; 1.
DR Pfam; PF00723; Glyco_hydro_15; 1.
DR PIRSF; PIRSF01031; Glu-a-glcsd_SBD; 1.
DR PRINTS; PRO0736; GLHYDRLASE15.
DR PROSITE; PS00820; GLUCOAMYLASE; 1.
DR Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal.
FT SIGNAL. 1 19 POTENTIAL.
FT PROPEP 20 35 POTENTIAL.
FT CHAIN 36 626 GLUCOAMYLASE.
FT BINDING 155 155 SUBSTRATE (BY SIMILARITY).
FT ACT_SITE 211 211 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 214 214 GENERAL ACID CATALYST (BY SIMILARITY).
FT ACT_SITE 215 215 INTERACT WITH SUBSTRATE (BY SIMILARITY).
FT CARBOHYD 106 106 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 206 206 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 82 82 MISSING (IN REF. 1).
FT CONFLICT 550 550 A -> R (IN REF. 1).
FT CONFLICT 560 560 V -> L (IN REF. 1).
FT SEQUENCE 626 AA; 66474 MW; 5485BDBBA7A3E349 CRC64;
Query Match 15.2%; Score 73; DB 1; Length 626;
Best Local Similarity 26.9%; Pred. No. 4.8;
Matches 29; Conservative 19; Mismatches 34; Indels 26; Gaps 6;
QY 4 SLTVGAGVASADPVDVAVITTCNYG----QVVAALNATDPGAAQF---NASPVAQSYL- 55
DB 392 SVSTGYSSSTSYTAIIINAVTYADGVFDVIAQYTPDGSIAEGFDKDGAPLSATHLT 451
QY 56 ----RNFLAA-----PPQRAAQAQVPGAAQYIGLVESVAGS 92
DB 452 WSVASFLSAARRAGIVPSWGAASA---NSLPGSCS----ASTVAGS 492
RESULT 4
YD33_MYCTU STANDARD; PRT; 344 AA.
AC Q10674;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein RV1333/MT1375/Mb1368.
GN RV1333 OR MT1375 OR MTCY130.18 OR MB1368.
OS Mycobacterium tuberculosis, and
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence."
RN Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayan L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL J. Bacteriol. 184:5479-5490(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Akin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: STRONG TO M.LEPRAE ML1167.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
DR EMBL; Z73902; CAA98097.1; -.
DR EMBL; AE007011; AAK45639.1; -.
DR EMBL; BX248338; CAD94229.1; -.
DR PIR; A70771; A70771.
DR MEROPS; T04.001; -.
DR TIGR; MT1375; -.
DR Tuberculist; RV1333; -.
DR InterPro; IPR005321; Peptidase_S58.
DR Pfam; PF03576; Peptidase_S58; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 25 45 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.
FT TRANSMEM 104 124 POTENTIAL.
FT TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 161 181- POTENTIAL.
FT TRANSMEM 224 244 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 302 322 POTENTIAL.
FT SEQUENCE 344 AA; 33952 MW; 296A055FECF8E650 CRC64;
Query Match 15.0%; Score 72; DB 1; Length 344;
Best Local Similarity 34.1%; Pred. No. 3.2;
Matches 29; Conservative 7; Mismatches 29; Indels 20; Gaps 3;
QY 2 AMSLTVGAGVASAD-----PVDVAVITTCNYGQVVAALNATDPGAAQNASPVAQSY 54
DB 155 AGALKGGVGTASATLQSGVTGVLAVN-----AAGNVVDPATGLPKNADLVGEFA 205
QY 55 LRNFLLAAPPPQRAAQAQVPGCA 79
DB 206 LR----APPAEQIAALAQSLSPGLA 226
RESULT 5
VCAA_BPT7
ID VCAA_BPT7 STANDARD; PRT; 345 AA.
AC P19726; P03177;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Major capsid protein 10A.
GN 10.

```


DR PRINTS; PRO0368; FADPDR.
DR PRINTS; PRO0945; HGRDTASE.
DR PRINTS; PRO0411; FNDRTASEI.
DR PRODOM; PD000139; FAD_Pyr_redox; 1.
DR PROSITE; PS01047; HMA 1; 1.
DR PROSITE; PS0846; HMA 2; 1.
DR PROSITE; PS00076; PYRIDINE REDOX 1; 1.
KW Mercuric resistance; Oxidoreductase; Flavoprotein; FAD; NADP;
KW Mercury; Redox-active center; Metal-binding; Plasmid.
FT DOMAIN 1 66 HMA.
FT NP_BIND 128 136 FAD (ADP PART) (BY SIMILARITY).
FT DISULFID 136 141 REDOX-ACTIVE.
FT NON_CONS 342 343
FT METAL 457 457 MERCURY (POTENTIAL).
FT METAL 458 458 MERCURY (POTENTIAL).
SQ SEQUENCE 460 AA; 48587 MW; 6284666EF3F653F05 CRC64;

Query Match 14.3%; Score 69; DB 1; Length 460;
Best Local Similarity 32.6%; Pred. No. 9.1;
Matches 29; Conservative 14; Mismatches 30; Indels 16; Gaps 5;

QY 9 AGVASADPVDAVNTTCNYGVVAALNATDPGAAQFNASPVQAQSVLRNFLAAPPQRAA 68
Db 64 AMLADAPPTD---NRTGLDFKVRGWMGAADKSGAE---RPLQ-----VAVIGSGGA 110

QY 69 MAAQLQVPAQAQYIGLVE--SVAGSCNN 95
Db 111 MAAALKAVEGQAQ-VTLIERGTIGTCVN 138

RESULT 8
REGA RHOSU
ID REGA RHOSU STANDARD; PRT; 183 AA.
AC O82868;
DT 30-MAY-2000 (Rel. 39, Created)
DT 20-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Photosynthetic apparatus regulatory protein regA.
GN REGA.
OS Rhodovulum sulfidophilum (Rhodobacter sulfidophilus).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Rhodovulum.
OX NCBI_TaxID=35806;
RN [1]
RP SEQUENCE FROM N.A.
RA Masuda S.;
RT "Rhodovulum sulfidophilum photosynthetic regulatory genes.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM REGB/REGA.
CC INVOLVED IN TRANSACTIVATING ANAEROBIC EXPRESSION OF THE
CC PHOTOSYNTHETIC APPARATUS. IT IS A TRANSCRIPTIONAL REGULATOR THAT
CC IS RESPONSIBLE FOR ACTIVATING EXPRESSION OF THE PUF, PUFH, AND PUC
CC OPERONS IN RESPONSE TO A DECREASE IN OXYGEN TENSION (BY
CC SIMILARITY).
CC -!- PTM: Phosphorylated by regB (Probable).
CC -!- SIMILARITY: Contains 1 response regulatory domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB010722; BAA31474.1; -.
DR HSSP; P10958; 1DBW.
DR InterPro; IPR002197; HTH Fis.
DR InterPro; IPR008329; Res_reg_RegA.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF00072; response_reg; 1.
DR PIRSF; PIRSF036385; Res_reg_RegA; 1.
DR PRODOM; PD000039; Response_reg; 1.

DR SMART: SMO0448; REC; 1.
DR TIGRFAMs; TIGR01199; HTH fis; 1.
KW PROSITE; PS0110; RESPONSE REGULATORY; 1.
KW Sensory transduction; Phosphorylation; Transcription regulation;
KW DNA-binding; Activator.
FT DOMAIN 13 127 RESPONSE REGULATORY.
FT MOD_RES 62 62 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 183 AA; 20219 MW; 394DB79D39AAB23F CRC64;

Query Match 14.1%; Score 68; DB 1; Length 183;
Best Local Similarity 27.7%; Pred. No. 4.3;
Matches 28; Conservative 9; Mismatches 32; Indels 32; Gaps 3;

QY 3 MSLTVGAG--VASADP-----
Db 40 MAETVAAGKAIASRPPAYAVVDLEDTGLDVVETLREKRPDAKIVLTGYGAIATAV 99

QY 35 NATDPGA-----AAQFNASPVQAQSVLRNFLAAPPQRAAMAA 71
Db 100 AAVKVGATDYLKPADANDVTAALLSNGEALPPPPENPMMSA 140

RESULT 9
SRIA PHYPO
ID SRIA PHYPO STANDARD; PRT; 246 AA.
AC P09350;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Spherulin 1A precursor.
OS Physarum polycephalum (Slime mold).
OC Eukaryota; Mycetozoa; Myxogastria; Myxogastromycetidae; Physariida;
OC Physarum.
OX NCBI_TaxID=5791;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=88137962; PubMed=2830170;
RT Bernier F., Lemieux G., Pallotta D.;
RT "Gene families encode the major encystment-specific proteins of
RT Physarum polycephalum plasmodia.";
RL Gene 59:265-277(1987).
RN [2]
RP PARTIAL SEQUENCE OF 20-34.
RA MEDLINE=93049354; PubMed=1425703;
RA Lane B.G., Cuming A.C., Fregeau J., Carpita N.C., Hurkman W.J.,
RA Bernier F., Dratewka-Kos E., Kennedy T.D.;
RT "Germin isoforms are discrete temporal markers of wheat development.
RT pseudogermin is a uniquely thermostable water-soluble oligomeric
RT protein in ungerminated embryos and like germin in germinated
RT embryos, it is incorporated into cell walls.";
RL Eur. J. Biochem. 209:961-969(1992).
CC -!- SUBCELLULAR LOCATION: Cell wall.
CC -!- DEVELOPMENTAL STAGE: Accumulates specifically during spherulation.
CC -!- MISCELLANEOUS: Spherulin is a major encystment-specific protein.
CC -!- SIMILARITY: Belongs to the germin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M18428; AAA29982.1; -.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin sup.
DR InterPro; IPR001929; Germin.
DR Pfam; PF00190; Cupin; 1.
DR PRINTS; PR00325; GERMIN.
DR PROSITE; PS00725; GERMIN; 1.
KW Cell wall; Signal; Glycoprotein; Manganese; Metal-binding;
KW Multigene family.

FT SIGNAL 1 19
 FT CHAIN 20 245
 FT METAL 123 123
 FT METAL 125 125
 FT METAL 130 130
 FT METAL 170 170
 FT CARBOHYD 213 213
 SQ SEQUENCE 246 AA; A06360A64ALB8BD2 CRC64;
 Query Match 14.0%; Score 67.5; DB 1; Length 246;
 Best Local Similarity 30.3%; Pred. No. 6.7; Indels 15; Gaps 3;
 Matches 25; Conservative 8; Mismatches 34; Indels 15; Gaps 3;
 QY 6 TVGAGVASADPVDV---INTTCNYGVVAALNATDPG---AAQFNASP-----V 50
 DB 155 TLEAGMATVPGCAIHFEINNCPEAMFVAFFNEDPGVQTASSFFGLPADVGVSLNI 214
 QY 51 AQSLRNFLAAPPQRAAAQ 72
 DB 215 SSIQTVEDLGKLPONFAVAMQ 236

RESULT 10
 CPN_DROME STANDARD; PRT; 865 AA.
 AC Q02910;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Calphotin.
 GN CPN OR CAP.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=Canton-S;
 RX MEDLINE=93165729; PubMed=8094559;
 RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
 RT "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
 [2]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=Canton-S;
 RX MEDLINE=93165730; PubMed=8434015;
 RA Ballinger D.G., Xue N., Harshman K.D.;
 RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
 calcium and contains a leucine zipper";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
 CC
 CC -!- FUNCTION: Might function as a calcium-sequestering "sponge" to
 regulate the amount of free cytoplasmic calcium. It binds 0.3 mole
 of Ca(2+) per mole of protein.
 CC
 CC -!- SUBUNIT: Homodimer (Probable).
 CC
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; hypodense compartment.
 CC
 CC -!- TISSUE SPECIFICITY: Soma and axons of photoreceptor cells of
 compound eyes and ocelli.
 CC
 CC -!- DEVELOPMENTAL STAGE: Expressed early in photoreceptor cell
 development.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L02111; AAA28405.1; -;
 CC DR EMBL; L05080; AAA28420.1; -;
 CC DR EMBL; A47282; A47282.
 CC DR EMBL; A47283; A47283.

DR FlyBase; FBgn0010218; Cpn.
 DR GO; GO:000509; F:calcium ion binding; IDA.
 KW Calcium-binding.
 FT CONFLICT 36 36 A -> AVAPAVVA (IN REF. 2).
 FT CONFLICT 43 43 I -> T (IN REF. 2).
 FT CONFLICT 64 64 I -> V (IN REF. 2).
 FT CONFLICT 76 76 T -> A (IN REF. 2).
 FT CONFLICT 100 100 P -> PP (IN REF. 2).
 FT CONFLICT 126 127 VQ -> AP (IN REF. 2).
 FT CONFLICT 154 154 I -> V (IN REF. 2).
 FT CONFLICT 160 160 S -> T (IN REF. 2).
 FT CONFLICT 534 534 A -> E (IN REF. 2).
 FT CONFLICT 699 699 I -> T (IN REF. 2).
 FT CONFLICT 703 703 V -> L (IN REF. 2).
 FT CONFLICT 721 721 D -> E (IN REF. 2).
 SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CFE CRC64;
 Query Match 14.0%; Score 67.5; DB 1; Length 865;
 Best Local Similarity 25.1%; Pred. No. 25; Indels 9; Gaps 2;
 Matches 23; Conservative 10; Mismatches 37; Indels 9; Gaps 2;
 QY 16 PVDVINTTCNYGVVAALNATDPGAAAFNAPVQAGYLNFIAAPPQRAAAQAQQA 75
 DB 9 PVSAPVAAPVTPSAVAPVQVWSPAAVAPAPAPAPIATVPV-----APPPTLASVQPATVT 63
 QY 76 VPG-----AAQYIGLVESVA 90
 DB 64 IPAPAPIAAASVTPVASVA 82
 RESULT 11
 FXD1_HUMAN STANDARD; PRT; 465 AA.
 ID FXD1_HUMAN
 AC Q16676; Q12949;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Forkhead box protein D1 (Forkhead-related protein FKHL8) (Forkhead-
 related transcription factor 4) (FREAC-4).
 GN FOXD1 OR FKHL8 OR FREAC4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96355467; PubMed=8702877;
 RA Ernstsson S., Pierrou S., Hulander M., Cederberg A., Hellqvist M.,
 RA Carlsson P., Enerback S.;
 RT "Characterization of the human forkhead gene FREAC-4. Evidence for
 regulation by Wilms' tumor suppressor gene (WT-1) and p53.";
 RL J. Biol. Chem. 271:21094-21099(1996).
 [2]
 RN
 RP SEQUENCE OF 120-225 FROM N.A.
 RX MEDLINE=95045392; PubMed=7957066;
 RA Pierrou S., Hellqvist M., Samuelsson L., Enerbaeck S., Carlsson P.;
 RT "Cloning and characterization of seven human forkhead proteins:
 binding site specificity and DNA bending.";
 RL EMO J. 13:5002-5012(1994).
 CC
 CC -!- FUNCTION: Binding of FREAC-3 and FREAC-4 to their cognate sites
 results in bending of the DNA at an angle of 80-90 degrees.
 CC
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC
 CC -!- SIMILARITY: Contains 1 fork-head domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U59832; AAC50661.1; -;
 CC DR EMBL; U59832; AAC50661.1; -;

DR EMBL: U59831; AAC50660.1; -;
 DR EMBL: U13222; AAA92039.1; -;
 DR PIR: G02738; G02738.
 DR PIR: S51627; S51627.
 DR HSP: O63245; 2HPH.
 DR TRANSFAC: T02472; -;
 DR Genew: HGNC:3802; FOXD1.
 DR MIM: 601091; -;
 DR GO: GO:0003700; Fitranscription factor activity; TAS.
 DR InterPro: IPR001766; TF_Fork_head.
 DR Pfam: PF00250; Fork_head; 1.
 DR PRINTS: PRO0053; FORKHEAD.
 DR ProDom: PD000425; TF_Fork_head; 1.
 DR SMART: SM00339; FH; 1.
 DR PROSITE: PS00657; FORK_HEAD_1; 1.
 DR PROSITE: PS00658; FORK_HEAD_2; 1.
 DR PROSITE: PS00309; FORK_HEAD_3; 1.
 KW Transcription regulation; DNA-binding; Nuclear protein.
 FT DOMAIN 26 34 POLY-GLU.
 FT DOMAIN 39 43 POLY-GLY.
 FT DOMAIN 52 57 POLY-ARG.
 FT DOMAIN 69 72 POLY-GLU.
 FT DOMAIN 73 76 POLY-ASP.
 FT DOMAIN 97 113 POLY-GLY.
 FT DNA_BIND 124 215 FORK-HEAD.
 FT DOMAIN 231 234 POLY-ALA.
 FT DOMAIN 252 256 POLY-ALA.
 FT DOMAIN 259 266 POLY-PRO.
 FT DOMAIN 293 303 POLY-ALA.
 FT DOMAIN 309 315 POLY-PRO.
 FT DOMAIN 375 390 POLY-ALA.
 FT DOMAIN 428 434 POLY-ALA.
 SQ SEQUENCE 465 AA; 46140 MW; D3E7854909CCBF AE CRC64;
 Query Match 13.9%; Score 67; DB 1; Length 465;
 Best Local Similarity 29.0%; Pred. No. 15;
 Matches 27; Conservative 15; Mismatches 37; Indels 14; Gaps 4;
 QY 8 GAGVASADP--VDVINTTCNVGVVVAALNTPGAAQAFNAPVAQSVLNFLLAAPPPQ 65
 Db 353 GASALARSFSTESITGG--SLGPAAAAAQAQAQAQAQASP-----SPVAAPAP 403
 QY 66 RAA--MAAQLOAVPGAAQYIGLVESVAGSCNN 95
 Db 404 GSSGGCAQAAGVGPAAALTRSLVAAAAAASS 436
 RESULT 12
 SCWA_YEAST
 ID SCWA_YEAST STANDARD; PRT; 389 AA.
 AC Q04951;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable family 17 glucosidase SCW10 precursor (EC 3.2.1.-) (Soluble
 cell wall protein 10).
 GN SCW10 OR YMR305C OR YM9952.07C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RX MEDLINE=97313268; PubMed=9169872;
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
 RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrrell B.G.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
 XII.";
 RL Nature 387:90-93(1997).
 RN [2]

RP SEQUENCE OF 30-39, AND SUBCELLULAR LOCATION.
 RC STRAIN=SEY6210;
 RX MEDLINE=98422453; PubMed=9748433;
 RA Cappellaro C., Mrsa V., Tanner W.;
 RT "New potential cell wall glucanases of Saccharomyces cerevisiae and
 RT their involvement in mating";
 RL J. Bacteriol. 180:5030-5037(1998).
 CC -!- FUNCTION: Glucanases possibly play a role in cell expansion
 CC during growth, in cell-cell fusion during mating, and in spore
 CC release during sporulation (by similarity).
 CC -!- SUBCELLULAR LOCATION: Cell wall.
 CC -!- PTM: Glycosylated.
 CC -!- SIMILARITY: Belongs to family 17 of glycosyl hydrolases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: Z49212; CAA89138.1; -;
 DR PIR: S53975; S53975.
 DR GerMOnline; 142983; -;
 DR SGD: S0004921; SCW10.
 DR GO: GO:0009277; C:cell wall (sensu Fungi); IDA.
 DR GO: GO:0005783; C:endoplasmic reticulum; IDA.
 DR GO: GO:0000747; P:conjugation with cellular fusion; IGI.
 DR InterPro: IPR000490; Glyco_hydro_17.
 DR Pfam: PF00332; Glyco_hydro_17; 1.
 DR PROSITE: PS00587; GLYCOSYL_HYDROL_F17; FALSE NEG.
 KW Hydrolase; Glycosidase; Cell wall; Glycoprotein; Signal.
 FT SIGNAL 1 18
 FT PROPEP 19 29
 FT CHAIN 30 389 PROBABLE FAMILY 17 GLUCOSIDASE SCW10.
 FT DOMAIN 65 134 SER-RICH.
 FT ACT_SITE 326 326 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 380 380 PROTON DONOR (BY SIMILARITY).
 FT CATHED 279 279 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 38 39 AQ -> QE (IN REF. 2).
 SQ SEQUENCE 389 AA; 40469 MW; 21F394BD41137DE1 CRC64;
 Query Match 13.8%; Score 66.5; DB 1; Length 389;
 Best Local Similarity 27.4%; Pred. No. 14;
 Matches 31; Conservative 12; Mismatches 49; Indels 21; Gaps 6;
 QY 2 AMSLTVCAG----VASADPVDVINT--TCNYGVVVAALNA---TDPGAAAFNAPVAQ 52
 Db 249 AKSALTSGAGTGVSVSVDTFTFIANNPDLNCSYSDYM-ANNAHAYFDENTAAQ-DAGPWVL 306
 QY 53 SYLNFLLAAPPPQRAAQAOL-----QAAPGAAQYIGLVESVAGSCNN 95
 Db 307 EQIERVVTACGGKDVITETGHPKGDYGEAVPSKANQEAIAISSIKSSCGS 359
 RESULT 13
 TSSP_HUMAN
 ID TSSP_HUMAN STANDARD; PRT; 514 AA.
 AC Q9NQE7; Q7541.6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Thymus-specific serine protease precursor (EC 3.4.-.-).
 GN PRSS16 OR TSSP.
 OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Thymus;
 RX MEDLINE=99458809; PubMed=10527559;

```

RA Bowlus C.L., Ahn J., Chu T., Gruen J.R.;
RT "Cloning of a novel MHC-encoded serine peptidase highly expressed by
RT cortical epithelial cells of the thymus.";
RL Cell. Immunol. 196;80-86(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Phillips S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Protease that may play a role in T-cell development.
CC -!- SUBCELLULAR LOCATION: Vesicular, either lysosomal or endosomal.
CC -!- TISSUE SPECIFICITY: Expressed predominantly in cortical thymic
CC epithelial cells.
CC -!- DEVELOPMENTAL STAGE: Expressed in fetal thymus.
CC -!- SIMILARITY: Belongs to peptidase family S28.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF052514; AAC33563.1; -
CC EMBL; AL021808; CAB94769.1; ALT_SEQ.
CC MEROPS; S28.003; -.
CC Genew; HGNC:9480; PRSS16.
CC MIM; 607169; -.
CC GO; GO:0008236; F:serine-type peptidase activity; NAS.
CC GO; GO:0030163; P:protein catabolism; NAS.
CC InterPro; IPR008758; Peptidase_S28.
CC Pfam; PF05577; Peptidase_S28; 1.
CC Hydrolase; Serine protease; Signal.
FT SIGNAL 1 24
FT CHAIN 25 514 THYMUS-SPECIFIC SERINE PROTEASE.
FT ACT_SITE 185 185 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 447 447 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 472 472 CHARGE RELAY SYSTEM (POTENTIAL).
FT CARBOHYD 70 70 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 321 321 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SEQUENCE 514 AA; 55048 MW; EF39BC06C560709A CRC64;

Query Match 13.8%; Score 66.5; DB 1; Length 514;
Best Local Similarity 31.8%; Pred.No.19;
Matches 31; Conservative 14; Mismatches 30; Indels 23; Gaps 5;

QY 9 AGVASADPDVDVAVINTTCYGVVVA-ALNATDFGAAQAQFNAS-----PVAQ 52
Db 204 ASVASAPVRAVLDFS-EYNDVVSRLMSTALGSLSCRAAVSAPAEVERLRSGAAQ 262
QY 53 SYLENFLAAPP-----QRAAMAQLOA-VPGNAQYIG 84
Db 263 AALRTLSACGPIGRAENQAELGALQALVGGVVQYDG 300

RESULT 14
DCOS_ECOLI
ID -DCOS_ECOLI STANDARD; PRT; 732 AA.
AC P24169;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ornithine decarboxylase, inducible (EC 4.1.1.17).
GN SPEG OR B0693.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
CX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92041955; PubMed=1939141;
RA Kashiwagi K., Suzuki T., Furuchi T., Kobayashi H.,

```

```

RA Igarashi K.;
RT "Coexistence of the genes for putrescine transport protein and
RT ornithine decarboxylase at 16 min on Escherichia coli chromosome.";
RL J. Biol. Chem. 266:20922-20927(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R.; Plunkett G. III; Bloch C.A.; Perna N.T.; Burland V.;
RA Riley M.; Collado-Vides J.; Glasner J.D.; Rode C.K.; Mayhew G.F.;
RA Gregor J.; Davis N.W.; Kirkpatrick H.A.; Goeden M.A.; Rose D.J.;
RA Mau B.; Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayaishi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
CC -!- CATALYTIC ACTIVITY: L-ornithine = putrescine + CO(2).
CC -!- COFACTOR: Pyridoxal phosphate.
CC -!- PATHWAY: Biosynthesis of spermidine from arginine; third step.
CC -!- INDUCTION: By low environmental pH.
CC -!- SIMILARITY: Belongs to the Orn/Lys/Arg decarboxylase class-I
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M64495; AAA62785.1; -
CC EMBL; AE000172; AAC73787.1; -
CC EMBL; D90708; BAA35349.1; -
CC PIR; A40839; A40839.
CC HSSP; P43099; 1C4K.
CC EcGene; EG10964; speF.
CC InterPro; IPR000310; Decarboxylase.
CC InterPro; IPR008286; Decarboxylase_C.
CC InterPro; IPR005308; OKR_DC_1_N.
CC Pfam; PF01276; OKR_DC_1_1.
CC Pfam; PF03711; OKR_DC_1_C; 1.
CC Pfam; PF03709; OKR_DC_1_N; 1.
CC PROSITE; PS00703; OKR_DC_1; 1.
CC Spermidine biosynthesis; Lyase; Decarboxylase; Pyridoxal phosphate;
CC Complete proteome.
CC BINDING 355 355 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC SEQUENCE 732 AA; 82415 MW; C2C030BEAB9D23F6 CRC64;

Query Match 13.8%; Score 66.5; DB 1; Length 732;
Best Local Similarity 26.0%; Pred.No.27;
Matches 34; Conservative 12; Mismatches 38; Indels 47; Gaps 5;

QY 9 AGVASADPDVDVAVINTTCYGVVVAALNA-----TDPGAAQAQFNASPV- 50
Db 184 AKVENADKTYFLVNGTSSNKKV--LNALLTPGDLVLFDRNHNKHHGALLQAGATPVY 241
QY 51 -----AQSYLNFLEAPPQRAAMAQLO-AVPGAQYIG----- 84
Db 242 LETARNPYFGIGDAHCPEESYLRELIAEVAPQRAKEAPFRLAVIQLGTGYDGTYNAR 301
QY 85 -LVESVAGSCN 94

```

Db 302 QVVDKIGHLCD 312

RESULT 15

VGLH HSVBC

ID VGLH HSVBC STANDARD; PRT; 842 AA.

AC P27539;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 16-OCT-2001 (Rel. 46, Last annotation update)

DE Glycoprotein H precursor.

GN GH OR UL22.

OS Bovine herpesvirus type 1 (strain Cooper).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Alphaherpesvirinae; Varicellovirus.

OX NCBI_TaxID=10323;

RN [1]

RP SEQUENCE FROM N.A.

RA Meyer A.L., Petrovskis E.A., Duffus W.P.N., Thomsen D.R.,

RA Post L.E.;

RL Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: GH AND GL FORM A COMPLEX THAT INDUCES NEUTRALIZING AND

CC SYNCYTIA INHIBITING ANTIBODIES (BY SIMILARITY).

CC -!- SIMILARITY: Belongs to the herpesviruses glycoprotein H family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X58867; CAA41677.1; -.

DR EMBL; Z78205; CAB01604.1; -.

DR EMBL; AJ004801; CAA06112.1; -.

DR PIR; S18462; S18462.

KW Glycoprotein; Transmembrane; Signal.

FT SIGNAL 1 25 POTENTIAL.

FT CHAIN 26 842 GLYCOPROTEIN H.

FT CARBOHYD 77 77 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 617 617 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 666 666 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 760 760 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 783 783 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 842 AA; 88375 MW; BA90759A74715F98 CRC64;

Query Match 13.8%; Score 66.5; DB 1; Length 842;

Best Local Similarity 28.4%; Pred. No. 31;

Matches 25; Conservative 11; Mismatches 43; Indels 9; Gaps 3;

QY 15 DPVDAVINTCNYG--QVVAALNATDGAAGFNA---SPVAQSYLRNFLAAPPORA 67

Db 447 EPAGAALESVVARGRKRAVAFSGPDAPLAARLLAHGVVSDLYDAFLRGELTWGPFNRH 506

QY 68 AM--AAQLQAVPGAQYIGLVESVAGSC 93

Db 507 ALFFFAVAASAFPADAQALELARDVTRKC 534

Search completed: June 22, 2004, 17:17:00

Job time : 3.68859 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2004, 16:55:54 ; Search time 7.84949 Seconds
(without alignments)
3958.816 Million cell updates/sec

Title: US-09-886-349A-22

Perfect score: 481
Sequence: 1 VAMSLTVGAGVASADPVDV.....PCAAQYIGLVESVAGSCNNY 96

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	481	100.0	110	050430	050430 mycobacteri
2	481	100.0	110	0700G8	Q700G8 mycobacteri
3	79	16.4	1243	Q9W6T6	Q9W6T6 brachydanio
4	78	16.2	280	Q8S2D6	Q8S2D6 oryza sativ
5	77	16.0	410	Q8RZ79	Q8RZ79 oryza sativ
6	75.5	15.7	336	Q98HE8	Q98HE8 rhizobium 1
7	75	15.6	337	Q8NTY7	Q8NTY7 corynebacte
8	74	15.4	586	Q8LN25	Q8LN25 oryza sativ
9	74	15.4	2213	Q8XN25	Q8XN25 corynebacte
10	73.5	15.3	462	Q8VPA4	Q8VPA4 pseudomonas
11	73.5	15.3	466	Q8DWE7	Q8DWE7 streptococc
12	73.5	15.3	1489	Q8DWE5	Q8DWE5 agrobacteri
13	72.5	15.1	190	Q84JP1	Q84JP1 arabidopsis
14	72.5	15.1	197	Q9S9P9	Q9S9P9 arabidopsis
15	72.5	15.1	654	Q8P938	Q8P938 xanthomonas
16	72	15.0	143	Q83VK1	Q83VK1 sphingomonas

17	72	15.0	196	16	Q9JW92	Q9JW92 neisseria m
18	72	15.0	241	16	Q9K3I3	Q9K3I3 streptomyc
19	71.5	14.9	308	12	Q39491	Q39491 bovine herp
20	71	14.8	314	5	Q8INP3	Q8INP3 drosophila
21	71	14.8	372	5	Q8INP4	Q8INP4 drosophila
22	71	14.8	441	16	Q8XQP8	Q8XQP8 raietonia s
23	71	14.8	516	16	Q7UPE9	Q7UPE9 rhodopirell
24	71	14.8	849	12	Q9WQY5	Q9WQY5 bovine herp
25	70.5	14.7	222	2	Q93L09	Q93L09 burkholderi
26	70.5	14.7	249	5	Q9UIN3	Q9UIN3 cryptospori
27	70.5	14.7	462	16	Q9HWK6	Q9HWK6 pseudomonas
28	70.5	14.7	555	5	P90625	P90625 cryptospori
29	70.5	14.7	618	11	Q91221	Q91221 mus musculu
30	70.5	14.7	689	11	Q9DBI6	Q9DBI6 mus musculu
31	70.5	14.7	689	11	Q8C2Q3	Q8C2Q3 mus musculu
32	70.5	14.7	1832	5	O96503	O96503 cryptospori
33	70.5	14.7	1850	5	Q9GRJ7	Q9GRJ7 leishmania
34	70	14.6	179	5	Q9VW18	Q9VW18 drosophila
35	70	14.6	196	16	Q9JXN1	Q9JXN1 neisseria m
36	70	14.6	293	5	Q17222	Q17222 caenorhabdi
37	70	14.6	411	16	Q8G431	Q8G431 bifidobacte
38	70	14.6	414	16	Q8PPX6	Q8PPX6 xanthomonas
39	70	14.6	876	4	Q8WY43	Q8WY43 homo sapien
40	70	14.6	1064	4	Q9HAN2	Q9HAN2 homo sapien
41	70	14.6	7525	2	Q9KIE0	Q9KIE0 streptomyc
42	69.5	14.4	273	16	Q7UKL8	Q7UKL8 rhodopirell
43	69.5	14.4	288	17	Q9HNP1	Q9HNP1 halobacteri
44	69.5	14.4	728	16	Q8XQ78	Q8XQ78 raietonia s
45	69.5	14.4	759	10	Q94KS5	Q94KS5 hordeum vul

ALIGNMENTS

RESULT 1

050430	PRELIMINARY;	PRT;	110 AA.
ID	O50430		
AC	O50430;		
DT	01-JUN-1998 (TREMBLrel. 06; Created)		
DT	01-JUN-1998 (TREMBLrel. 06; Last sequence update)		
DT	01-JUN-2003 (TREMBLrel. 24; Last annotation update)		
DE	Hypothetical protein Rv1174c.		
GN	Rv1174c OR MTV005.10C OR MT1211.		
OS	Mycobacterium tuberculosis.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.		
OX	NCBI_TaxID=1773;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=H37RV;		
RX	MEDLINE=98295987; PubMed=9634230;		
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,		
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,		
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,		
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,		
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,		
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,		
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,		
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;		
RT	"Deciphering the biology of Mycobacterium tuberculosis from the		
RT	complete genome sequence."		
RL	Nature 393:537-544(1998).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CDC 1551 / Oshkosh;		
RA	Fleischman R.D., Alland D., Eisen J.A., Carpenter L., White O.,		
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,		
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,		
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,		
RA	Bishai W.;		
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and		
RT	laboratory strains."		
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		

```

DR EMBL; AL010186; CAA15851.1; -.
DR EMBL; AE006998; AAK45468.1; -.
DR PIR; F70875; F70875.
DR TIGR; MT1211; -.
DR Tuberculist; Fv1174c; -.
DR KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 110 AA; 10881 MW; 0309DDB489150061 CRC64;

Query Match 100.0%; Score 481; DB 16; Length 110;
Best Local Similarity 100.0%; Pred. No. 3.1e-39;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGGAAQFNASPVASQYLNFLA 60
Dy 15 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGGAAQFNASPVASQYLNFLA 74
Qy 61 APPQRAAMAQAQVAGAAQYIGLVESVAGSCNNY 96
Dy 75 APPQRAAMAQAQVAGAAQYIGLVESVAGSCNNY 110

RESULT 2
Q7U0G8 PRELIMINARY; PRT; 110 AA.
AC Q7U0G8;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Hypothetical protein.
GN MB1207C.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett R., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248337; CAD94068.1; -.
KW Complete proteome.
SQ SEQUENCE 110 AA; 10881 MW; 0309DDB489150061 CRC64;

Query Match 100.0%; Score 481; DB 16; Length 110;
Best Local Similarity 100.0%; Pred. No. 3.1e-39;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGGAAQFNASPVASQYLNFLA 60
Dy 15 VAMSLTVGAGVASADPVDVAVINTTCNYGVVAALNATDGGAAQFNASPVASQYLNFLA 74
Qy 61 APPQRAAMAQAQVAGAAQYIGLVESVAGSCNNY 96
Dy 75 APPQRAAMAQAQVAGAAQYIGLVESVAGSCNNY 110

RESULT 3
Q9W6T6 PRELIMINARY; PRT; 1243 AA.
AC Q9W6T6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Patched-2 protein.
GN PFC2 OR PFC1.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;

```

```

OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Lewis K.E., Concordet J.P., Ingham P.W.;
RT "Characterisation of the second ptc gene in zebrafish.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ007742; CAB39726.1; -.
DR ZFIN; ZDB-GENE-980526-196; ptc2.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0008158; P:hedghog receptor activity; IEA.
DR InterPro; IPR003392; Patched.
DR InterPro; IPR004766; Patched_rec.
DR InterPro; IPR000731; SSD_STM.
DR Pfam; PF02460; Patched; 1.
DR TIGRFAMs; TIGR00918; 2A060602; 1.
DR PROSITE; PS00156; SSD; 1.
SQ SEQUENCE 1243 AA; 138548 MW; FAF459DDB0C91371 CRC64;

Query Match 16.4%; Score 79; DB 13; Length 1243;
Best Local Similarity 29.3%; Pred. No. 49;
Matches 27; Conservative 11; Mismatches 40; Indels 14; Gaps 3;

Qy 18 DAVINTTCNYGVVAALNATDP--GAAQFNASPVASQYLR-----NFLAAPP 64
Dy 925 DGIINPNAFY-IYLSNVSDPVAAYASQANIRHPPEWLHRTDSIPASRLNIPAAEPI 983
Qy 65 QRAAMAAQQAQVAGAAQYIGLVESVAGSCNNY 96
Dy 984 EYAQFPFYINGLRETPOFVEAIESVRAICNNY 1015

RESULT 4
Q8S2D6 PRELIMINARY; PRT; 280 AA.
AC Q8S2D6;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE P0401G10.4 protein (Similar to RNA-binding protein).
GN P0401G10.4 OR Q1294_F06.21.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC
clone P0401G10.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
clone OJ1294_F05.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003238; BAB88997.1; -.
DR EMBL; AP004326; BAB92880.1; -.
DR Gramene; Q8S2D6; -.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR PROSITE; PS00102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 280 AA; 28689 MW; 852687706D3DBC5F CRC64;

Query Match 16.2%; Score 78; DB 10; Length 280;
Best Local Similarity 34.2%; Pred. No. 12;
Matches 27; Conservative 6; Mismatches 42; Indels 4; Gaps 2;

```


QY 9 AGVASADPV-DVAVINTTCNGVVAALNATDPGAAQFNASPVQAQSYLRFNLAAPPORA 67
 DB 138 AGLTWANTGGGLSPFYQYGPVAVANAAAGYAGMQY---PQNYQYVAAAAAAGATAA 244
 QY 68 AMAAQLOAVPQAAQYIGLV 86
 DB 245 TTASQLTAVAGLQVQVAGTV 263

RESULT 5
 Q8R279 PRELIMINARY; PRT; 410 AA.
 ID Q8R279
 AC Q8R279
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE B1065G12.16 protein.
 GN B1065G12.16
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, BAC
 clone:B1065G12.";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBSJ databases.
 CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
 DR EMBL; AP003791; BAB90534.1; -.
 DR Gramene; Q8R279; -.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR InterPro; IPR002018; CarbohydraseB.
 DR InterPro; IPR000379; Ser_estrs.
 DR Pfam; PF00135; Coesterase; 1.
 KW Hydrolase.
 SQ SEQUENCE 410 AA; 44077 MW; 64F2F2C2B5E1D3B7 CRC64;

Query Match 16.0%; Score 77; DB 10; Length 410;
 Best Local Similarity 28.3%; Pred. No. 22;
 Matches 28; Conservative 12; Mismatches 49; Indels 10; Gaps 3;

QY 4 SLTVGAGVASADPVD--AVINTTCNGVQVVA---ALNATDPGAAQFNASPVQAQSYLRFN 57
 DB 23 SLAAAAAAAAGGEDDAAGFVSCRADEATAPSNPAFSAADGVASKDLHIDENSSLSVRI 82
 QY 58 FLAAPP-----QRAAMAAQLQAVPQAAQYIGLVESVAGS 92
 DB 83 FLUPTPPPHSRRASEPPPPATANGSGAPYRGYLDPHAVSS 121

RESULT 6
 Q98HE8 PRELIMINARY; PRT; 336 AA.
 ID Q98HE8
 AC Q98HE8
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Flagellar motor switch protein, Flig.
 GN MLI2901.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214958;
 RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,

RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
 Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP003000; BAB49918.1; -.
 DR GO; GO:000288; C:flagellum (sensu Bacteria); IEA.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0005840; C:ribosome; IEA.
 DR GO; GO:0003774; F:motor activity; IEA.
 DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
 DR GO; GO:0005935; P:chemotaxis; IEA.
 DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.
 DR GO; GO:0006412; P:protein biosynthesis; IEA.
 DR InterPro; IPR000090; Flg_Motor_Flig.
 DR InterPro; IPR000307; Ribosomal_S16.
 DR Pfam; PF01706; Flig-C; 1.
 DR PRINTS; PR00954; FLGMOTORFLIG.
 DR PROSITE; PS00732; RIBOSOMAL_S16; 1.
 DR PROSITE; PS00732; RIBOSOMAL_S16; 1.
 KW Complete proteome.
 SQ SEQUENCE 336 AA; 35986 MW; 90FCFA6ED7B0FEDB CRC64;

Query Match 15.7%; Score 75.5; DB 16; Length 336;
 Best Local Similarity 30.6%; Pred. No. 25;
 Matches 26; Conservative 13; Mismatches 39; Indels 7; Gaps 2;

QY 6 TVGAGVA-SADPVDVAVINTTCNGVQVVAALNATDPGAAQF-----NASPVQAQSYLRFN 58
 DB 67 TEGAGLSDSDRMDTILNLSLSPESAIMGNKKPEAAPGPPPIWPDLEKLEPSRLGTF 126
 QY 59 LAAPPPORAAMAAQLQAVPQAAQYI 83
 DB 127 LAGEHPQTRAMVLSKLAPQAAASVL 151

RESULT 7
 Q8NTY7 PRELIMINARY; PRT; 337 AA.
 ID Q8NTY7
 AC Q8NTY7
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Predicted dehydrogenases and related proteins (EC 1.1.1.18).
 GN CGL0164.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
 OX NCBI_TaxID=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S.;
 RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AP005274; BAB97557.1; -.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR000683; GFO_IDH_MocA.
 DR InterPro; IPR004104; GFO_IDH_MocA_C.
 DR Pfam; PF01408; GFO_IDH_MocA; 1.
 DR Pfam; PF02894; GFO_IDH_MocA_C; 1.
 KW Oxidoreductase; Complete proteome.
 SQ SEQUENCE 337 AA; 36431 MW; F5EF6EAF6A65924A3 CRC64;

Query Match 15.6%; Score 75; DB 16; Length 337;
 Best Local Similarity 37.3%; Pred. No. 28;
 Matches 28; Conservative 11; Mismatches 26; Indels 10; Gaps 4;

QY 7 VGAGVASADPVDVAVINTTCNGVQVVAALNATDPG-AAQFNASPVQAQSYLRFNLAAPP 63
 DB 10 VGAGAGAGADHIDRINNRTS--GAHISAIIEPDAARAAAAEDAPGAQAFTRIEDAIAAD- 66

65 AAPPAARVARAAPLAPKPGTFLQGV 90

RESULT 11

Q8DWE7 PRELIMINARY; PRT; 466 AA.

AC Q8DWE7

DT 01-MAR-2003 (TREMELrel. 23, Created)

DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Putative PTS system, fructose-specific IIBC component.

GN SMU114.

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1309;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=UA159 / ATCC 700610 / Serotype C;

RX MEDLINE=2229563; PubMed=12397186;

RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J., Qian Y., Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.; "Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen.";

RT Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).

RL EMBL; AB014863; AAN57895.1; --

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005351; F:sugar porter activity; IEA.

DR GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. .; IEA.

DR InterPro; IPR003352; Ptrans_ELIC.

DR InterPro; IPR003353; Ptrans_IIB_fruc.

DR InterPro; IPR006327; PTS_IIC_funct.

DR Pfam; PF02378; PTS_ELIC; 1.

DR Pfam; PF02379; PTS_IIB_fruc; 1.

DR TIGRFAMs; TIGR00829; FRU; 1.

DR TIGRFAMs; TIGR01427; PTS_IIC_fructo; 1.

KW Complete proteome.

SQ SEQUENCE 466 AA; 48902 MW; 3C7244E4B7A95815 CRC64;

Query Match 15.3%; Score 73.5; DB 16; Length 466;

Best Local Similarity 26.0%; Pred. No. 56;

Matches 25; Conservative 16; Mismatches 46; Indels 9; Gaps 2;

QY 1 VAMSLTVGAGVASADPVDVAVINTTCNGVQVVAALNATDFGAAQFNASPVASQYLRFNFA 60

DB 150 VLVAISFLGIYFDPKSYQYH-----FAALLKTDIGGVAMGWTILAAAYAESIA 202

QY 61 APPPPRAAQAQVPGAAQYIGLVESVAGSCNNY 96

DB 203 KRPGFVAGFVGGMVANGSGFLGGI--TAGFAAGY 236

RESULT 12

Q8U9P5 PRELIMINARY; PRT; 1489 AA.

AC Q8U9P5

DT 01-JUN-2002 (TREMELrel. 21, Created)

DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Polyketide synthetase.

GN ATU3681 OR AGR_L_2313.

OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.

OX NCBI_TaxID=176299;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21608550; PubMed=11743193;

RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P., Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L., Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr., Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,

Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A., Raymond C., Rouse C., Saenphimmachak C., Wu Z., Romero P., Gordon D., Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M., Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-X., Dolan M., Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V., Nester E.W.;

RT "The genome of the natural genetic engineer Agrobacterium tumefaciens C58.";

RT Science 294:2317-2323 (2001).

RN [2]

RX SEQUENCE FROM N.A.

RX MEDLINE=21608551; PubMed=11743194;

RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L., Hummel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F., Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B., Flanagan C., Crowell C., Gursen J., Lono C., Sear C., Strub G., Cielo C., Slater S.;

RT "Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58.";

RT Science 294:2323-2328 (2001).

DR EMBL; AE009297; AAL44493.1; ALT_INIT.

DR EMBL; AE008316; AAK89721.1; --

DR PIR; AG3009; AG3009.

DR PIR; G98274; G98274.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.

DR GO; GO:0008152; P:metabolism; IEA.

DR InterPro; IPR001227; AC:trans.

DR InterPro; IPR00794; Ketoacyl synth.

DR InterPro; IPR006162; P:antennae_S.

DR InterPro; IPR006163; P:bind.

DR Pfam; PF00698; Acyl transf; 1.

DR Pfam; PF00109; ketoacyl-synt; 1.

DR Pfam; PF02801; ketoacyl-synt; 1.

DR Pfam; PF00550; PP-binding; 1.

DR PROSITE; PS00675; ACP DOMAIN; 1.

DR PROSITE; PS00606; B_KETOACYL-SYNTHASE; 1.

DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.

KW Complete proteome.

SQ SEQUENCE 1489 AA; 159121 MW; 5BF4FF9FARFC90009 CRC64;

Query Match 15.3%; Score 73.5; DB 16; Length 1489;

Best Local Similarity 28.6%; Pred. No. 2.1e+02;

Matches 26; Conservative 18; Mismatches 36; Indels 11; Gaps 4;

QY 5 LTVGAGVASADPVDVAVINTTCNGVQVVAALNATDFGAAQFNASPVASQYLRFNFA 61

DB 373 LDTAAGIA-----GFIKTVLCKHGRIPASLHFRQANAKIDFPASFPANVTVLRDWTAL 426

QY 62 PFPORAAQAQVPGAAQYIGLVESVAGS 92

DB 427 ESPRRAGISS--LGIGGTNAHVLEAPAGA 455

RESULT 13

Q84JF1 PRELIMINARY; PRT; 190 AA.

AC Q84JF1

DT 01-JUN-2003 (TREMELrel. 24, Created)

DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Putative transcription factor.

GN AtUG30500.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,

DR EMBL; AC009917; AAF19754.1; -.
DR PIR; A86430; A86430.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001289; TF_CBEF.
DR Pfam; PF02045; CBEF_NFYA; 1.
DR PRINTS; PR00616; CCAATSUBUNTB.
DR ProDom; PD003860; TF_CBEF; 1.
DR SMART; SM00521; CBF; 1.
SQ SEQUENCE 197 AA, 22127 MW; EB67C5050BE0F874 CRC64;

Query Match 15.1%; Score 72.5; DB 10; Length 197;
Best Local Similarity 31.0%; Pred. No. 27;
Matches 26; Conservative 8; Mismatches 23; Indels 27; Gaps 96

QY 26 NYGQVVAAALNATDPG-----AAAFNAPSVQASVLYRFLAAPPDQ-----RAAKRAQLQ 74
DB 41 NYEIVTSLVYSDPGTTNSMAPGQF---PYDPFYRS-IFAPDPQYTGVLQLMGVQQQ 96

QY 75 AVP-----GAAQYIGLV 86
DB 97 GVPLESDAVEPVPVFNQYHGIL 120

RESULT 15
Q8P938 PRELIMINARY; PRT; 654 AA.
ID Q8P938 AC Q8P938;
DT C1-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein XCC2028.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=3340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorllo C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.W., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Clapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Fomighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.K., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Subetal J.C., Kitajima J.P.;
R "Comparison of the genomes of two Xanthomonas pathogens with differing
R host specificities.";
RRL Nature 417:459-463(2002).
RRL EMBL; AE012308; ANA41317.1; -.
DR GO; GO:0004563; F:beta-N-acetylhexosaminidase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001540; Glyco_hydro_20.
DR Pfam; PF02839; Glyco_hydro_20b; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 654 AA; 69164 MW; 9A96E86E1302F410 CRC64;

Query Match 15.1%; Score 72.5; DB 16; Length 654;
Best Local Similarity 29.9%; Pred. No. 1e+02;
Matches 20; Conservative 11; Mismatches 21; Indels 15; Gaps 1

Db	588	VSAALANSQATQOFDAA-----KQLAEQAAAIQSIPGATRFAGPVKVA	632
Qy	90	AGSCNY	96
Db	633	DGVLSF	639

Search completed: June 22, 2004, 17:22:57
Job time : 10.8495 secs

100

100

100

100

100

100

100

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2004, 17:09:52 ; Search time 55 Seconds
(without alignments)
498.311 Million cell updates/sec

Title: US-09-886-349A-24

Perfect score: 468

Sequence: 1 MSLDAHPQLVQSAPAA.....EAGTYVAADAAASTYTG 97

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A Genesep29Jan04: *
1: Genesep1980s: *
2: Genesep1990s: *
3: Genesep2000s: *
4: Genesep2001s: *
5: Genesep2002s: *
6: Genesep2003as: *
7: Genesep2003bs: *
8: Genesep2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	468	100.0	97	2	Aaw73653 M. tuberc
2	468	100.0	97	2	Aaw73759 M. tuberc
3	468	100.0	97	2	Aaw73763 M. tuberc
4	468	100.0	97	2	Aaw73869 M. tuberc
5	468	100.0	97	4	Aab35221 M. tubercu
6	468	100.0	97	4	Aau08229 Mycobacte
7	468	100.0	97	4	Aau08209 Mycobacte
8	468	100.0	97	4	Aau01893 M. tuberc
9	468	100.0	97	5	Aae29712 Mycobacte
10	468	100.0	97	5	Aae17576 Mycobacte
11	468	100.0	299	5	Aay32067 Mycobacte
12	468	100.0	299	5	Aae29723 Mycobacte
13	468	100.0	299	5	Aae17587 Mycobacte
14	468	100.0	306	5	Aau74596 Antigenic
15	468	100.0	433	5	Aay32065 Mycobacte
16	468	100.0	433	5	Aau74594 Antigenic
17	468	100.0	710	5	Aay32066 Mycobacte
18	468	100.0	710	5	Aae29724 Mycobacte
19	468	100.0	710	5	Aae17588 Mycobacte
20	468	100.0	723	5	Aau74595 Antigenic
21	468	100.0	856	2	Aay32064 Mycobacte
22	468	100.0	859	5	Aau74593 Antigenic
23	442	94.4	97	4	Aab35228 M. tubercu
24	352	75.2	98	5	Abu05649 M. tuberc
25	90.5	19.3	104	7	Adel0630 Structura

26	89.5	19.1	405	6	ABU34103	Abu34103 Protein e
27	88.5	18.9	163	6	ABU34018	Abu34018 Protein e
28	87.5	18.7	95	2	AAW32376	AAW32376 Mycobacte
29	87.5	18.7	95	2	AAW32444	AAW32444 Mycobacte
30	87.5	18.7	95	2	AAW64321	AAW64321 Mycobacte
31	87.5	18.7	95	2	AAW81747	AAW81747 M. tuberc
32	87.5	18.7	95	2	AAW32097	AAW32097 Mycobacte
33	87.5	18.7	95	2	AAW38981	AAW38981 M. tuberc
34	87.5	18.7	95	2	AAW39118	AAW39118 M. tuberc
35	87.5	18.7	95	5	AAE29717	AAE29717 Mycobacte
36	87.5	18.7	95	5	AAE17581	AAE17581 Mycobacte
37	87.5	18.7	100	2	AAW64339	AAW64339 Mycobacte
38	87.5	18.7	100	2	AAW81706	AAW81706 M. tuberc
39	87.5	18.7	100	2	AAW03705	AAW03705 M. tuberc
40	87.5	18.7	100	2	AAW38993	AAW38993 M. tuberc
41	87.5	18.7	100	2	AAW39136	AAW39136 M. tuberc
42	87.5	18.7	100	4	AAW19845	AAW19845 Mycobacte
43	87.5	18.7	100	4	AAW35218	AAW35218 M. tubercu
44	87.5	18.7	100	5	ABG30969	ABG30969 Mycobacte
45	87.5	18.7	100	5	ABU05988	ABU05988 M. tuberc

ALIGNMENTS

RESULT 1
AAW73653
ID AAW73653 standard; protein; 97 AA.
XX
AC AAW73653;
XX
XX 24-MAR-1999 (first entry)
XX
DE M. tuberculosis antigen clone Tb472 protein sequence.
XX
XX Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
KW infection.
XX
XX Mycobacterium tuberculosis.
OS
XX
XX W09853076-A2.
XX
XX 26-NOV-1998.
XX
XX 20-MAY-1998; 98WO-US010514.
XX
XX 20-MAY-1997; 97US-00858998.
PR 05-MAY-1998; 98US-00073009.
XX
XX (CORI-) CORIXA CORP.
XX
XX Alderson MR, Dillon DC, Skeiky YAW, Campos-Neto A;
XX
XX WPI; 1999-045315/04.
XX
XX N-PSDB; AAX01132.
XX
XX New isolated Mycobacterium tuberculosis antigens - used to develop
XX products for the prevention, treatment and diagnosis of tuberculosis
XX infection.
XX
XX Claim 2; Page 76-77; 104pp; English.

XX This sequence represents an immunogenic portion of a Mycobacterium
XX tuberculosis antigen of the invention. The polypeptides are useful for
XX immunotherapy to treat or prevent tuberculosis (especially in humans),
XX e.g. they can be included with an acceptable carrier in pharmaceutical
XX compositions or included in vaccines, and administered to induce
XX protective immunity in a patient against M. tuberculosis. Tuberculosis is
XX a chronic, infectious disease generally caused by M. tuberculosis
XX infection, and if left untreated typically results in serious
XX complications and death. Fusion proteins containing the antigen, or DNA
XX molecules can similarly be included with an acceptable carrier in
XX pharmaceutical compositions or in vaccines and administered as above. The

CC polypeptides are also useful for diagnosis of tuberculosis, by contacting
 CC dermal cells with at least one polypeptide and detecting an immune
 CC response (especially induration) on the patient's skin. Inhibiting the
 CC spread of tuberculosis requires vaccination and accurate diagnosis, since
 CC antibiotic therapy may not be effective due to the existence of an
 CC asymptomatic but contagious stage and to patient non-compliance. The
 CC polypeptides overcome concerns of safety and efficacy of current
 CC vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
 CC lack of sensitivity and specificity of existing diagnostic techniques
 XX
 XX Sequence 97 AA;

Query Match 100.0%; Score 468; DB 2; Length 97;
 Best Local Similarity 100.0%; Pred. No. 2.3e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAFQAAWSAQAFHQGESSAAFOAAHARFVA 60
 DB 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAFQAAWSAQAFHQGESSAAFOAAHARFVA 60
 QY 61 AAQVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97
 DB 61 AAQVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97

RESULT 2
 AAW73759
 ID AAW73759 standard; peptide; 97 AA.
 XX
 AC AAW73759;
 XX
 DT 24-MAR-1999 (first entry)
 DE
 DE M. tuberculosis antigen clone protein #1.
 XX
 KW Antigen; M. tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
 KW infection.
 XX Mycobacterium tuberculosis.
 XX WO9853076-A2.
 XX 26-NOV-1998.
 XX 20-MAY-1998; 98WO-US010514.
 XX 20-MAY-1997; 97US-00858998.
 PR 05-MAY-1998; 98US-00073009.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 PI Alderson MR, Dillon DC, Skeiky YAW, Campos-Neto A;
 XX
 DR WPI; 1999-045315/04.
 XX
 XX New isolated Mycobacterium tuberculosis antigens - used to develop
 PT products for the prevention, treatment and diagnosis of tuberculosis
 PT infection.
 XX
 XX Disclosure; Page 92; 104pp; English.

XX This sequence represents an immunogenic portion of a Mycobacterium
 CC tuberculosis antigen of the invention. The polypeptides are useful for
 CC immunotherapy to treat or prevent tuberculosis (especially in humans),
 CC e.g. they can be included with an acceptable carrier in pharmaceutical
 CC compositions or included in vaccines, and administered to induce
 CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
 CC a chronic, infectious disease generally caused by M. tuberculosis
 CC infection, and if left untreated typically results in serious
 CC complications and death. Fusion proteins containing the antigen, or DNA
 CC molecules can similarly be included with an acceptable carrier in
 CC pharmaceutical compositions or in vaccines and administered as above. The
 CC polypeptides are also useful for diagnosis of tuberculosis, by contacting

CC dermal cells with at least one polypeptide and detecting an immune
 CC response (especially induration) on the patient's skin. Inhibiting the
 CC spread of tuberculosis requires vaccination and accurate diagnosis, since
 CC antibiotic therapy may not be effective due to the existence of an
 CC asymptomatic but contagious stage and to patient non-compliance. The
 CC polypeptides overcome concerns of safety and efficacy of current
 CC vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
 CC lack of sensitivity and specificity of existing diagnostic techniques
 XX
 XX Sequence 97 AA;

Query Match 100.0%; Score 468; DB 2; Length 97;
 Best Local Similarity 100.0%; Pred. No. 2.3e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAFQAAWSAQAFHQGESSAAFOAAHARFVA 60
 DB 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAFQAAWSAQAFHQGESSAAFOAAHARFVA 60
 QY 61 AAQVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97
 DB 61 AAQVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97

RESULT 3
 AAW73763
 ID AAW73763 standard; protein; 97 AA.
 XX
 AC AAW73763;
 XX
 DT 24-MAR-1999 (first entry)
 DE
 DE M. tuberculosis antigen clone Tb472 protein sequence.
 XX
 KW Antigen; M. tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
 KW infection.
 XX Mycobacterium tuberculosis.
 XX WO9853075-A2.
 XX 26-NOV-1998.
 XX 20-MAY-1998; 98WO-US010407.
 XX 20-MAY-1997; 97US-00859381.
 PR 05-MAY-1998; 98US-00073010.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 PI Alderson MR, Dillon DC, Skeiky YAW, Campos-Neto A;
 XX
 DR WPI; 1999-045314/04.
 DR N-PSDB; AAX01166.
 XX
 XX Polypeptide comprising immunogenic Mycobacterium tuberculosis antigen -
 PT useful for immunisation against M. tuberculosis infection to treat or
 PT prevent tuberculosis, and in diagnosis of tuberculosis.
 XX
 XX Claim 2; Page 75-76; 100pp; English.

XX This sequence represents an immunogenic portion of a Mycobacterium
 CC tuberculosis antigen of the invention. The polypeptides are useful for
 CC immunotherapy to treat or prevent tuberculosis (especially in humans),
 CC e.g. they can be included with an acceptable carrier in pharmaceutical
 CC compositions or included in vaccines, and administered to induce
 CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
 CC a chronic, infectious disease generally caused by M. tuberculosis
 CC infection, and if left untreated typically results in serious
 CC complications and death. Fusion proteins containing the antigen, or DNA
 CC molecules can similarly be included with an acceptable carrier in
 CC pharmaceutical compositions or in vaccines and administered as above. The
 CC polypeptides are also useful for diagnosis of tuberculosis, by contacting

CC dermal cells with at least one polypeptide and detecting an immune
CC response (especially induration) on the patient's skin. Inhibiting the
CC spread of tuberculosis requires vaccination and accurate diagnosis, since
CC antibiotic therapy may not be effective due to the existence of an
CC asymptomatic but contagious stage and to patient non-compliance. The
CC polypeptides overcome concerns of safety and efficacy of current
CC vaccination with live bacteria (usually *Bacillus Calmette-Guerin*) and
CC lack of sensitivity and specificity of existing diagnostic techniques
XX
SQ Sequence 97 AA;

Query Match 100.0%; Score 468; DB 2; Length 97;
Best Local Similarity 100.0%; Pred. No. 2.3e-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMGAQAFHQGESSAAAFQAAHARFVA 60
DB 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMGAQAFHQGESSAAAFQAAHARFVA 60

OY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAASTYTGF 97
DB 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAASTYTGF 97

RESULT 4
AAW73869
ID AAW73869 standard; peptide; 97 AA.
AC AAW73869;
XX
XX
DT 24-MAR-1999 (first entry)
DE M. tuberculosis antigen cloneb protein #1.
XX
XX Antigen; M. tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
KW infection.
XX
OS Mycobacterium tuberculosis.
XX
XX WO9853075-A2.
XX
PD 26-NOV-1998.
XX
XX 20-MAY-1998; 98WO-US010407.
XX
PR 20-MAY-1997; 97US-00859381.
PR 05-MAY-1998; 98US-00073010.
XX
XX (CORI-) CORIXA CORP.
PA
PI Alderson MR, Dillon DC, Skeiky YAW, Campos-Neto A;
XX
XX WPI; 1999-045314/04.
DR
XX
PT Polypeptide comprising immunogenic Mycobacterium tuberculosis antigen -
PT useful for immunisation against M. tuberculosis infection to treat or
PT prevent tuberculosis, and in diagnosis of tuberculosis.
XX
PS Disclosure; Page 91; 100pp; English.
XX
XX

This sequence represents an immunogenic portion of a Mycobacterium
CC tuberculosis antigen of the invention. The polypeptides are useful for
CC immunotherapy to treat or prevent tuberculosis (especially in humans),
CC e.g. they can be included with an acceptable carrier in pharmaceutical
CC compositions or included in vaccines, and administered to induce
CC protective immunity in a patient against M. tuberculosis. Tuberculosis is
CC a chronic, infectious disease generally caused by M. tuberculosis
CC infection, and if left untreated typically results in serious
CC complications and death. Fusion proteins containing the antigen, or DNA
CC molecules can similarly be included with an acceptable carrier in
CC pharmaceutical compositions or in vaccines and administered as above. The
CC polypeptides are also useful for diagnosis of tuberculosis, by contacting
CC dermal cells with at least one polypeptide and detecting an immune

CC response (especially induration) on the patient's skin. Inhibiting the
CC spread of tuberculosis requires vaccination and accurate diagnosis, since
CC antibiotic therapy may not be effective due to the existence of an
CC asymptomatic but contagious stage and to patient non-compliance. The
CC polypeptides overcome concerns of safety and efficacy of current
CC vaccination with live bacteria (usually *Bacillus Calmette-Guerin*) and
CC lack of sensitivity and specificity of existing diagnostic techniques
XX
SQ Sequence 97 AA;

Query Match 100.0%; Score 468; DB 2; Length 97;
Best Local Similarity 100.0%; Pred. No. 2.3e-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMGAQAFHQGESSAAAFQAAHARFVA 60
DB 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMGAQAFHQGESSAAAFQAAHARFVA 60

OY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAASTYTGF 97
DB 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAASTYTGF 97

RESULT 5
AAB35221
ID AAB35221 standard; protein; 97 AA.
XX
XX AAB35221;
AC
XX
DT 24-APR-2001 (first entry)
DE M tuberculosis Rv0287 protein.
XX
XX Tuberculosis; TB; vaccine; esat-6 gene family; Rv0287; Rv1036c; Rv1037c;
KW Rv2346c; Rv2348c; Rv2853c; Rv2654c; Rv3020c; Rv3444c; Rv3445c; Rv3890c;
KW Rv3891c; Rv3904c; Rv3905c.
XX
OS Mycobacterium tuberculosis.
XX
XX WO200104151-A2.
XX
XX 18-JAN-2001.
XX
XX 13-JUL-2000; 2000WO-DX000398.
XX
XX 13-JUL-1999; 99DK-00001020.
PR 15-JUL-1999; 99US-0144011P.
XX
XX (STAT-) STATENS SERUM INST.
PA
XX
XX Andersen P, Skjot R;
PI
XX
XX WPI; 2001-091923/10.
DR
XX N-PSDB; AAF24411.
XX
PT New polypeptide encoded by a member of the esat-6 gene family for
PT immunizing against and diagnosis of tuberculosis.
XX
XX
PS Example 2; Page 67; 80pp; English.
XX
XX

The present invention provides the protein and coding sequences for
CC members of the esat-6 gene family from Mycobacterium tuberculosis. These
CC proteins include Rv0287, Rv1037c, Rv2346c, Rv2348c, Rv2653c,
CC Rv2654c, Rv3020c, Rv3444c, Rv3445c, Rv3890c, Rv3891c, Rv3904c and
CC Rv3905c. These can be used to produce vaccines against, and in the
CC diagnosis of, tuberculosis (TB) infection. The present sequence is one of
CC the proteins of the invention
XX
SQ Sequence 97 AA;

Query Match 100.0%; Score 468; DB 4; Length 97;
Best Local Similarity 100.0%; Pred. No. 2.3e-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAFOAAHARFVA 60
 Db 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAFOAAHARFVA 60
 QY 61 AAANKVNTLLDVAQANLGEAGTGYVAADAAAASITYTGF 97
 Db 61 AAANKVNTLLDVAQANLGEAGTGYVAADAAAASITYTGF 97

RESULT 6
 AAU08229
 ID AAU08229 standard; protein; 97 AA.
 AC AAU08229;
 XX
 DT 17-DEC-2001 (first entry)
 DE Mycobacterium tuberculosis polypeptide for MTB98..
 XX
 KW Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial;
 KW immunostimulant; MTB98.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO200162893-A2.
 XX
 PD 30-AUG-2001.
 XX
 PF 26-FEB-2001; 2001WO-US005992.
 XX
 PR 25-FEB-2000; 2000US-0185037P.
 PR 08-AUG-2000; 2000US-0223828P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Skeiky Y, Ovendale P, Jen S, Lodes M;
 DR WPI; 2001-536638/59.
 XX
 PT An isolated polypeptide comprising a Mycobacterium antigen, e.g., from
 PT Mycobacterium tuberculosis, useful in a vaccine for enhancing an immune
 PT response to and inhibiting development of a Mycobacterium infection.
 XX
 PS Disclosure; Page 160; 161pp; English.
 XX

CC The present invention relates to the isolation of Mycobacterium
 CC tuberculosis antigen polypeptides (e.g. Tb224) and the nucleic acids
 CC encoding them. The invention describes compounds and methods for the
 CC diagnosis of tuberculosis or for inducing protective immunity against
 CC tuberculosis. The compounds comprise at least one immunogenic portion of
 CC one or more Mycobacterium proteins and nucleic acid molecules encoding
 CC such polypeptides. The Mycobacterium proteins and nucleic acid molecules
 CC of the invention and antibodies directed against the Mycobacterium
 CC proteins may be used in vaccines for immunisation against Mycobacterium
 CC infections. The nucleic acids encoding the Mycobacterium proteins may be
 CC used in gene therapy. The present sequence represents the M. tuberculosis
 CC polypeptide for MTB9.8
 XX
 SQ Sequence 97 AA;

Query Match 100.0%; Score 468; DB 4; Length 97;
 Best Local Similarity 100.0%; Pred. No. 2.3e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAFOAAHARFVA 60
 Db 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAFOAAHARFVA 60
 QY 61 AAANKVNTLLDVAQANLGEAGTGYVAADAAAASITYTGF 97
 Db 61 AAANKVNTLLDVAQANLGEAGTGYVAADAAAASITYTGF 97

Db 61 AAANKVNTLLDVAQANLGEAGTGYVAADAAAASITYTGF 97

RESULT 7
 AAU08209
 ID AAU08209 standard; protein; 97 AA.
 AC AAU08209;
 XX
 DT 17-DEC-2001 (first entry)
 DE Mycobacterium tuberculosis antigen Tb472.
 XX
 KW Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial;
 KW immunostimulant; antigen Tb472.
 XX
 OS Mycobacterium tuberculosis.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 89..97
 FT /note= "Encoded by G"
 XX
 PN WO200162893-A2.
 XX
 PD 30-AUG-2001.
 XX
 PF 26-FEB-2001; 2001WO-US005992.
 XX
 PR 25-FEB-2000; 2000US-0185037P.
 PR 08-AUG-2000; 2000US-0223828P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Skeiky Y, Ovendale P, Jen S, Lodes M;
 DR WPI; 2001-536638/59.
 DR N-PSDB; AAS12467.
 XX
 PT An isolated polypeptide comprising a Mycobacterium antigen, e.g., from
 PT Mycobacterium tuberculosis, useful in a vaccine for enhancing an immune
 PT response to and inhibiting development of a Mycobacterium infection.
 XX
 PS Example 1; Page 145-146; 161pp; English.
 XX

CC The present invention relates to the isolation of Mycobacterium
 CC tuberculosis antigen polypeptides (e.g. Tb224) and the nucleic acids
 CC encoding them. The invention describes compounds and methods for the
 CC diagnosis of tuberculosis or for inducing protective immunity against
 CC tuberculosis. The compounds comprise at least one immunogenic portion of
 CC one or more Mycobacterium proteins and nucleic acid molecules encoding
 CC such polypeptides. The Mycobacterium proteins and nucleic acid molecules
 CC of the invention and antibodies directed against the Mycobacterium
 CC proteins may be used in vaccines for immunisation against Mycobacterium
 CC infections. The nucleic acids encoding the Mycobacterium proteins may be
 CC used in gene therapy. The present sequence represents M. tuberculosis
 CC antigen Tb472
 XX
 SQ Sequence 97 AA;

Query Match 100.0%; Score 468; DB 4; Length 97;
 Best Local Similarity 100.0%; Pred. No. 2.3e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAFOAAHARFVA 60
 Db 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAAPHQGESSAAFOAAHARFVA 60
 QY 61 AAANKVNTLLDVAQANLGEAGTGYVAADAAAASITYTGF 97
 Db 61 AAANKVNTLLDVAQANLGEAGTGYVAADAAAASITYTGF 97

CC heterologous antigens, as a fusion protein, and vectors expressing the
 CC fusion proteins are used as vaccines to prophylactically immunise mammals
 CC (especially humans) against infection by Mycobacteria. The compositions
 CC contain at least 2 heterologous antigens that increase the serological
 CC sensitivity of individuals infected with tuberculosis, a disease
 CC frequently affecting patients with acquired immunodeficiency disease,
 CC AIDS
 CC
 XX Sequence 97 AA;
 SQ
 Query Match 100.0%; Score 468; DB 4; Length 97;
 Best Local Similarity 100.0%; Pred. No. 2.3e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1 MSLLDHIPPOLVASQSAFAAKAGLMRHTIGQAEQAMSAQAQAFHQGESAAFOAAHARFVA 60
 |||||
 DB 1 MSLLDHIPPOLVASQSAFAAKAGLMRHTIGQAEQAMSAQAQAFHQGESAAFOAAHARFVA 60
 |||||
 QY 61 AAQKNTLLDVAQANLGEAAGTYVAADAAAATYTGTF 97
 |||||
 DB 61 AAQKNTLLDVAQANLGEAAGTYVAADAAAATYTGTF 97
 |||||
 RESULT 9
 AAE29712
 ID AAE29712 standard; protein; 97 AA.
 XX
 AC AAE29712;
 XX
 DT 27-JAN-2003 (first entry)
 XX
 DE Mycobacterium tuberculosis MSL antigenic protein.
 XX
 KW Vaccine; immunity; diagnostic agent; gene therapy; MSL antigen; MTB9.8.
 XX
 OS Mycobacterium tuberculosis.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 88..97
 FT /note= "Encoded by GCTG"
 XX
 PN WO200272792-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 13-MAR-2002; 2002WO-US008223.
 XX
 PR 13-MAR-2001; 2001US-0275837P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Brannon M, Guderian J;
 XX
 DR WI; 2002-759844/82.
 DR N-PSDB; AAD47088.
 XX
 PT New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 PT tuberculosis.
 XX
 PS Disclosure; Page 101; 155pp; English.
 XX
 CC The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as

RESULT 8
 AAU01893
 ID AAU01893 standard; protein; 97 AA.
 XX
 AC AAU01893;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE M. tuberculosis partial antigen Mtb9.8 (MSL).
 XX
 KW Mtb9.8; MSL; antigen; vaccine; tuberculosis; AIDS;
 KW acquired immunodeficiency disease.
 XX
 OS Mycobacterium tuberculosis.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..15
 FT /label= Mtb9.8_derived_peptide_1
 FT Peptide 6..20
 FT /label= Mtb9.8_derived_peptide_2
 FT Peptide 11..25
 FT /label= Mtb9.8_derived_peptide_3
 FT Peptide 14..30
 FT /label= Mtb9.8_derived_peptide_4
 FT Peptide 21..35
 FT /label= Mtb9.8_derived_peptide_5
 FT Peptide 26..40
 FT /label= Mtb9.8_derived_peptide_6
 FT Peptide 31..45
 FT /label= Mtb9.8_derived_peptide_7
 FT Peptide 36..50
 FT /label= Mtb9.8_derived_peptide_8
 FT Peptide 41..55
 FT /label= Mtb9.8_derived_peptide_9
 FT Peptide 46..60
 FT /label= Mtb9.8_derived_peptide_10
 FT Peptide 51..65
 FT /label= Mtb9.8_derived_peptide_11
 FT Peptide 56..70
 FT /label= Mtb9.8_derived_peptide_12
 FT Peptide 61..75
 FT /label= Mtb9.8_derived_peptide_13
 FT Peptide 66..80
 FT /label= Mtb9.8_derived_peptide_14
 FT Peptide 71..88
 FT /label= Mtb9.8_derived_peptide_15
 FT Misc-difference 88..97
 FT /note= "Encoded by GCTG"
 XX
 PN WO200124820-A1.
 XX
 PD 12-APR-2001.
 XX
 PF 10-OCT-2000; 2000WO-US028095.
 XX
 PR 07-OCT-1999; 99US-0158338P.
 PR 07-OCT-1999; 99US-0158425P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky Y, Reed S, Houghton RL, Mcneill PD, Dillon DC, Lodes ML;
 XX
 DR WI; 2001-290576/30.
 DR N-PSDB; AAS03784.
 XX
 PT Vaccinating against Mycobacteria infections in mammals using fusion
 PT proteins comprising combinations of heterologous antigens.
 XX
 PS Disclosure; Page 145; 168pp; English.
 XX
 CC The sequence represents Mycobacterium tuberculosis Mtb9.8 (also known as
 CC MSL), an M. tuberculosis antigen. Compositions comprising at least 2

CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is M.
 CC tuberculosis MSL antigenic protein. MSL is also referred to as MTB9.8
 XX
 SQ Sequence 97 AA;

Query Match 100.0%; Score 468; DB 5; Length 97;
 Best Local Similarity 100.0%; Pred. NO. 2.3e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSLDDAHIPQLVASQSAFAKAGLMRHTTIGQAEQAAMSAQAQFHGESSAAFAQAAHARFVA 60
 Db 1 MSLDDAHIPQLVASQSAFAKAGLMRHTTIGQAEQAAMSAQAQFHGESSAAFAQAAHARFVA 60

Qy 61 AAQVNTLLDVAQANLGEAAGTVAADAAAATYTG 97
 Db 61 AAQVNTLLDVAQANLGEAAGTVAADAAAATYTG 97

RESULT 10
 AAEL17576
 ID AAEL17576 standard; protein; 97 AA.
 XX
 AC AAEL17576;
 DT 22-APR-2002 (first entry)
 XX
 DE Mycobacterium tuberculosis MTB9.8 (MSL) protein.
 XX
 XX Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB9.8; MSL protein.
 XX
 OS Mycobacterium tuberculosis.
 XX
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 88.197
 FT /note= "Encoded by GCTC"
 XX
 XX WO200198460-A2.
 XX
 XX 27-DEC-2001.
 XX
 XX 20-JUN-2001; 2001WO-US019959.
 XX
 XX 20-JUN-2000; 2000US-00597796.
 PR 01-FEB-2001; 2001US-0265737P.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 XX Skeiky Y, Reed S, Alderson M;
 XX
 XX WPI; 2002-147798/19.
 DR N-PSDB; AAD28346.
 XX
 XX Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 XX Claim 9; Page 117; 136pp; English.
 XX
 XX The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the

CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is M. tuberculosis MTB9.8
 CC (MSL) protein
 XX
 SQ Sequence 97 AA;

Query Match 100.0%; Score 468; DB 5; Length 97;
 Best Local Similarity 100.0%; Pred. NO. 2.3e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSLDDAHIPQLVASQSAFAKAGLMRHTTIGQAEQAAMSAQAQFHGESSAAFAQAAHARFVA 60
 Db 1 MSLDDAHIPQLVASQSAFAKAGLMRHTTIGQAEQAAMSAQAQFHGESSAAFAQAAHARFVA 60

Qy 61 AAQVNTLLDVAQANLGEAAGTVAADAAAATYTG 97
 Db 61 AAQVNTLLDVAQANLGEAAGTVAADAAAATYTG 97

RESULT 11
 AAY32067
 ID AAY32067 standard; protein; 299 AA.
 XX
 AC AAY32067;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen fusion protein Mtb31f.
 XX
 KW Tuberculosis; antigen; fusion protein; Mtb31f; DPV; MSL; diagnosis;
 KW therapy; vaccine; immunogen.
 XX
 OS Mycobacterium tuberculosis.
 XX
 XX
 FH Key Location/Qualifiers
 FT Peptide 1.8
 FT /note= "Met/His tag"
 FT Protein 9.90
 FT /note= "DPV"
 FT Protein 93.186
 FT /note= "MTI"
 FT Protein 189.285
 FT /note= "MSL"
 XX
 XX WO9951748-A2.
 XX
 XX 14-OCT-1999.
 XX
 XX 07-APR-1999; 99WO-US007717.
 PR 07-APR-1998; 98US-00065556.
 PR 30-DEC-1998; 98US-00223040.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 XX Skeiky YAW, Alderson M, Campos-Neto A;
 XX
 XX WPI; 1999-601610/51.
 DR N-PSDB; AAZ20202.
 XX
 XX New fusion proteins useful for diagnosis, prevention and treatment of
 PT tuberculosis.
 XX
 XX Claim 1; Fig 9A-B; 83pp; English.
 XX
 XX This sequence represents a recombinant Mycobacterium tuberculosis tri-
 CC antigen fusion protein, termed Mtb31f, composed of the antigens DPV, MTI
 CC and MSL. The fusion protein is expressed in host cells using a vector
 CC carrying a polynucleotide (see AAZ20202) comprising the coding sequences

CC for the 3 antigens. The invention provides fusion proteins (see AAY32059-
CC 71) containing at least 2 M. tuberculosis antigens. The new fusion
CC proteins and polynucleotides encoding them are useful as vaccines for
CC preventing tuberculosis (claimed), for diagnosis (via in vitro assays or
CC intradermal skin tests for detection of anti-M. tuberculosis antibodies),
CC monitoring of disease progression, and treatment of tuberculosis. They
CC are more effective immunogens than mixtures of the individual protein
CC components
XX
XX Sequence 299 AA;

Query Match 100.0%; Score 468; DB 2; Length 299;
Best Local Similarity 100.0%; Pred. No. 9.6e-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSLLDAHIPOLVASOSAFKAGLMRHTTIGQAEQAMSAQAAPHQGESAAFOAHARFVA 60
Db 189 MSLLDAHIPOLVASOSAFKAGLMRHTTIGQAEQAMSAQAAPHQGESAAFOAHARFVA 248
QY 61 AAANKVTLDDVAQANLGEAAGTYVAADAAAATYTGTF 97
Db 249 AAANKVTLDDVAQANLGEAAGTYVAADAAAATYTGTF 285

RESULT 12
AAE29723
ID AAE29723 standard; protein; 299 AA.
XX
AC AAE29723;
DT 29-AUG-2003 (revised)
DT 27-JAN-2003 (first entry)
XX
DE Mycobacterium sp. DPV-MTI-MSL-MTCC#2 fusion protein.
KW Vaccine; immunity; diagnostic agent; gene therapy; DPV-MTI-MSL-MTCC#2;
KW MTB71F; MTB8.4; MTB9.9A; MTB9.8; MTB41; fusion protein.
XX
OS Mycobacterium tuberculosis.
OS Mycobacterium sp.
OS Chimeric.
XX
PN WO200272792-A2.
XX
PD 19-SEP-2002.
XX
PF 13-MAR-2002; 2002WO-US008223.
XX
PR 13-MAR-2001; 2001US-0275837P.
XX
PA (CORI-) CORIXA CORP.
PI Skeiky Y, Brannon M, Guderian J;
XX
DR WPI; 2002-759844/82.
DR N-PSDB; AAD47100.
XX
PT New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
PT tuberculosis.
XX
PS Disclosure; Page 114-115; 155pp; English.

XX The invention relates to a recombinant nucleic acid molecule encoding a
CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
CC polynucleotide sequence encoding an antigen or an antigenic fragment from
CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
CC polypeptide or its fragment. The Leishmania polynucleotide is selected
CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
CC are used in methods for eliciting immune response in mammals. They are
CC useful as vaccines to elicit protective immunity against pathogenic
CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion

CC polypeptides are used for enhancing the expression of polynucleotides, as
CC in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is
CC MTB71F fusion protein. This fusion protein comprises DPV (MTB8.4)-MTI
CC (MTB9.9A)-MSL (MTB9.8) protein sequences from M. tuberculosis and MTCC#2
CC (MTB41) protein from Mycobacterium sp. (Updated on 29-AUG-2003 to
CC standardise OS field)
XX
XX Sequence 299 AA;

Query Match 100.0%; Score 468; DB 5; Length 299;
Best Local Similarity 100.0%; Pred. No. 9.6e-48;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSLLDAHIPOLVASOSAFKAGLMRHTTIGQAEQAMSAQAAPHQGESAAFOAHARFVA 60
Db 189 MSLLDAHIPOLVASOSAFKAGLMRHTTIGQAEQAMSAQAAPHQGESAAFOAHARFVA 248
QY 61 AAANKVTLDDVAQANLGEAAGTYVAADAAAATYTGTF 97
Db 249 AAANKVTLDDVAQANLGEAAGTYVAADAAAATYTGTF 285

RESULT 13
AAE17587
ID AAE17587 standard; protein; 299 AA.
XX
AC AAE17587;
DT 22-APR-2002 (first entry)
DE Mycobacterium species MTB71F fusion protein.
KW Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; MTB71F; DPV-MTI-MSL-MTCC#2 protein.
XX
OS Mycobacterium sp.
XX WO200198460-A2.
XX
PD 27-DEC-2001.
XX
PF 20-JUN-2001; 2001WO-US019959.
XX
PR 20-JUN-2000; 2000US-00597796.
PR 01-FEB-2001; 2001US-0265737P.
XX
PA (CORI-) CORIXA CORP.
PI Skeiky Y, Reed S, Alderson M;
XX
DR WPI; 2002-147798/19.
DR N-ESDB; AAD28358.
XX
PT Composition comprising MTB39 antigen and MTB32A antigen from
PT Mycobacterium species, useful for eliciting immune response in a subject.
XX
PS Disclosure; Page 130-131; 136pp; English.

XX The present invention relates to fusion proteins containing at least two
CC Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected with
CC Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the

CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB71F (DPV-MTI-MSL-MTCC#2) fusion protein
 XX
 SQ Sequence 299 AA;

Query Match 100.0%; Score 468; DB 5; Length 299;
 Best Local Similarity 100.0%; Pred. No. 9.6e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAPFAKAGLMRHTIGQAEQAMSAQAFHQGESSAAFOAAHARFVA 60
 Db 189 MSLLDAHIPOLVASQSAPFAKAGLMRHTIGQAEQAMSAQAFHQGESSAAFOAAHARFVA 248
 QY 61 AAQVNTLLDVAQANLGEAAAGTYVAADAAASTYTGF 97
 Db 249 AAQVNTLLDVAQANLGEAAAGTYVAADAAASTYTGF 285

RESULT 14
 AAU74596
 ID AAU74596 standard; protein; 306 AA.

AC AAU74596;
 XX
 DT 29-AUG-2003 (revised)
 DT 08-MAY-2002 (first entry)
 XX
 DE Antigenic fusion protein DPV-MTI-MSL (Mtb31f).
 XX
 KW Fusion protein; tuberculosis; Mycobacterium tuberculosis;
 KW tuberculostatic; immunogen; vaccine; DPV-MTI-MSL; Mtb31f.
 XX
 OS Mycobacterium tuberculosis.
 OS Chimeric.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 300
 FT /label= OTHER
 FT /note= "OTHER= Xaa. Xaa= In frame stop codon"
 XX
 PN US2002009459-A1.
 XX
 PD 24-JAN-2002.
 XX
 PF 07-APR-1999; 99US-00287849.
 XX
 PR 13-MAR-1997; 97US-00818112.
 PR 01-OCT-1997; 97US-00942578.
 PR 18-FEB-1998; 98US-00025197.
 PR 07-APR-1998; 98US-00056556.
 PR 30-DEC-1998; 98US-00223040.
 XX
 XX (REED/) REED S G.
 PA (SKEI/) SKEIKY Y A.
 PA (DILL/) DILLON D C.
 PA (ALDE/) ALDERSON M.
 PA (CAMP/) CAMPOS-NETO A.

PI Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
 XX
 DR WPI; 2002-171134/22.
 DR N-PSDB; ABK14136.

XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for
 PT diagnosing, treating or preventing M. tuberculosis infection,
 PT particularly as vaccine for treating or preventing tuberculosis.
 XX
 PS Claim 1; Fig 9; 62pp; English.

XX The invention relates to a purified polypeptide which induces an immune
 CC response of Mycobacterium tuberculosis. Polypeptides of the invention are
 CC useful for diagnosing, treating or preventing M. tuberculosis infection,
 CC particularly tuberculosis infection. In particular, the polypeptides are
 CC useful as a vaccine formulation with an adjuvant to afford long-term
 CC protection in animals against the development of tuberculosis. The
 CC protein coding sequence may be used to encode a protein product for use
 CC as an immunogen to induce and/or enhance an immune response to M.
 CC tuberculosis. This sequence represents an M. tuberculosis fusion protein
 CC of the invention. (Updated on 29-AUG-2003 to standardise OS field)
 XX
 SQ Sequence 306 AA;

Query Match 100.0%; Score 468; DB 5; Length 306;
 Best Local Similarity 100.0%; Pred. No. 9.9e-48;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAPFAKAGLMRHTIGQAEQAMSAQAFHQGESSAAFOAAHARFVA 60
 Db 189 MSLLDAHIPOLVASQSAPFAKAGLMRHTIGQAEQAMSAQAFHQGESSAAFOAAHARFVA 248
 QY 61 AAQVNTLLDVAQANLGEAAAGTYVAADAAASTYTGF 97
 Db 249 AAQVNTLLDVAQANLGEAAAGTYVAADAAASTYTGF 285

RESULT 15
 AAY32065
 ID AAY32065 standard; protein; 433 AA.

AC AAY32065;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen fusion protein Mtb46f.
 XX
 KW Tuberculosis; antigen; fusion protein; Mtb46f; ERD14; DPV; MTI; MSL;
 KW diagnosis; therapy; vaccine; immunogen.

OS Mycobacterium tuberculosis.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..8
 FT Protein /note= "Met/His tag"
 FT Protein 9..152
 FT Protein /note= "ERD14"
 FT Protein 155..236
 FT Protein /note= "DPV"
 FT Protein 239..332
 FT Protein /note= "MTI"
 FT Protein 335..433
 FT Protein /note= "MSL"

WO9951748-A2.

PD 14-OCT-1999.
 XX
 PF 07-APR-1999; 99WO-US007717.
 XX
 PR 07-APR-1998; 98US-00056556.
 PR 30-DEC-1998; 98US-00223040.

XX (CORI-) CORIXA CORP.
 XX
 XX Skeiky YAW, Alderson M, Campos-Neto A;
 PI
 DR WPI; 1999-601610/51.
 DR N-PSDB; AAZ20200.
 XX
 PT New fusion proteins useful for diagnosis, prevention and treatment of
 PT tuberculosis.

PS Claim 1; Fig 7A-B; 83pp; English.

XX This sequence represents a recombinant Mycobacterium tuberculosis tetra-
CC antigen fusion protein, termed Mt46f, composed of the antigens ERD14,
CC DPV, M71 and MSL. The fusion protein is expressed in host cells using a
CC vector carrying a polynucleotide (see AAZ20200) comprising the coding
CC sequences for the 4 antigens. The invention provides fusion proteins (see
CC AAY32059-71) containing at least 2 M. tuberculosis antigens. The new
CC fusion proteins and polynucleotides encoding them are useful as vaccines
CC for preventing tuberculosis (claimed), for diagnosis (via in vitro assays
CC or intradermal skin tests for detection of anti-M. tuberculosis
CC antibodies), monitoring of disease progression, and treatment of
CC tuberculosis. They are more effective immunogens than mixtures of the
CC individual protein components

XX

SQ Sequence 433 AA;

Query Match 100.0%; Score 468; DB 2; Length 433;
Best Local Similarity 100.0%; Pred. No. 1.5e-47;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLDAHIPOLVASOSAFKAGLMRHTTIGQAEQAAMSAQAAPHQGESAAFOAHAREVA 60
|||
Db 335 MSLDAHIPOLVASOSAFKAGLMRHTTIGQAEQAAMSAQAAPHQGESAAFOAHAREVA 394

QY 61 AAKVNTLLDVAQANLGEAAGTYVAADAAAATYTCF 97
|||
Db 395 AAKVNTLLDVAQANLGEAAGTYVAADAAAATYTCF 431

Search completed: June 23, 2004, 17:12:30
Job time : 57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2004, 17:09:52 ; Search time 23 Seconds
(without alignments)
217.727 Million cell updates/sec

Title: US-09-886-349A-24
Perfect score: 468
Sequence: 1 MSLDHAHPQLVASQSAFAA.....EAGTYVADAAASSTYTG 97

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	468	100.0	97	4	US-09-073-009-109
2	468	100.0	97	4	US-09-073-009-143
3	468	100.0	97	4	US-09-073-010-109
4	468	100.0	97	4	US-09-073-010-143
5	468	100.0	299	4	US-09-287-849-19
6	468	100.0	433	4	US-09-287-849-14
7	468	100.0	710	4	US-09-287-849-16
8	468	100.0	856	4	US-09-287-849-12
9	87.5	18.7	95	3	US-08-818-112-88
10	87.5	18.7	95	4	US-08-818-111-89
11	87.5	18.7	95	4	US-09-056-556-88
12	87.5	18.7	95	4	US-09-072-596-89
13	87.5	18.7	95	4	US-09-072-967-88
14	87.5	18.7	100	3	US-08-818-112-115
15	87.5	18.7	100	4	US-09-056-556-115
16	87.5	18.7	100	4	US-09-116-492A-5
17	87.5	18.7	100	4	US-09-072-596-110
18	87.5	18.7	100	4	US-09-072-967-115
19	87.5	18.7	100	4	US-09-287-849-8
20	87.5	18.7	358	4	US-09-056-556-214
21	87.5	18.7	802	4	US-09-072-596-209
22	87.5	18.7	802	4	US-09-072-596-346
23	87.5	18.7	802	4	US-09-072-967-214
24	87.5	18.7	802	4	US-09-072-967-351
25	87.5	18.7	802	4	US-09-287-849-10
26	87.5	18.7	802	4	US-08-818-112-117
27	86.5	18.5	80	3	US-08-818-112-117

28	86.5	18.5	80	4	US-08-818-111-112	Sequence 112, App
29	86.5	18.5	80	4	US-09-056-556-117	Sequence 117, App
30	86.5	18.5	80	4	US-09-072-596-112	Sequence 112, App
31	86.5	18.5	80	4	US-09-072-967-117	Sequence 117, App
32	86	18.4	18	4	US-09-073-009-124	Sequence 124, App
33	86	18.4	18	4	US-09-073-010-124	Sequence 124, App
34	79.5	17.0	92	4	US-09-344-529-2	Sequence 2, Appli
35	79.5	17.0	361	4	US-09-543-681A-5390	Sequence 5390, Ap
36	77.5	16.6	400	4	US-09-073-009-126	Sequence 126, App
37	77.5	16.6	400	4	US-09-073-010-126	Sequence 126, App
38	77	16.5	15	4	US-09-073-009-118	Sequence 118, App
39	77	16.5	15	4	US-09-073-010-118	Sequence 118, App
40	76.5	16.3	286	4	US-09-489-039A-12707	Sequence 12707, A
41	75.5	16.1	469	4	US-09-489-039A-13565	Sequence 13565, A
42	75	16.0	15	4	US-09-073-009-114	Sequence 114, App
43	75	16.0	15	4	US-09-073-010-114	Sequence 114, App
44	75	16.0	339	1	US-08-433-854-4	Sequence 4, Appli
45	75	16.0	339	1	US-08-174-745A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-073-009-109
; Sequence 109, Application US/09073009
; Patent No. 6555653
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, David C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Columbia Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 05-MAY-1998
; APPLICATION NUMBER: US/09/073,009
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.441C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-073-009-109

TUBERCULOSIS AND MEI

Query Match 100.0%; Score 468; DB 4; Length 97;
Best Local Similarity 100.0%; Pred. No. 8.7e-53;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSLDAHPQLVASQSAFAAAGLMRHTTIGQAGAAASQAQAFHOGESSAAFAAHARFVA	60
DB	1	MSLDAHPQLVASQSAFAAAGLMRHTTIGQAGAAASQAQAFHOGESSAAFAAHARFVA	60

QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97
DB 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97

RESULT 2
US-09-073-009-143
; Sequence 143, Application US/09073009
; Patent No. 6555653
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Cumbria Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,009
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.441C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-073-009-143

TUBERCULOSIS AND ME

QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97
DB 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97

RESULT 3
US-09-073-010-109
; Sequence 109, Application US/09073010
; Patent No. 6613881
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS AND
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Cumbria Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,010
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.441C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-073-009-143

Query Match 100.0%; Score 468; DB 4; Length 97;
Best Local Similarity 100.0%; Pred. No. 8.7e-53;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDHIHPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAFOAAHARFVA 60
DB 1 MSLLDHIHPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAFOAAHARFVA 60

QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97
DB 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97

RESULT 4
US-09-073-010-143
; Sequence 143, Application US/09073010
; Patent No. 6613881
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS AND
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Cumbria Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,010
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Cumbria Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,010
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.440C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-073-010-109

TUBERCULOSIS AND ME

QY 1 MSLLDHIHPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAFOAAHARFVA 60
DB 1 MSLLDHIHPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAFOAAHARFVA 60
QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97
DB 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97

Query Match 100.0%; Score 468; DB 4; Length 97;
Best Local Similarity 100.0%; Pred. No. 8.7e-53;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDHIHPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAFOAAHARFVA 60
DB 1 MSLLDHIHPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAFOAAHARFVA 60

QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97
DB 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97

RESULT 4
US-09-073-010-143
; Sequence 143, Application US/09073010
; Patent No. 6613881
; GENERAL INFORMATION:
; APPLICANT: Alderson, Mark
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Campos-Neto, Antonio
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS AND
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY
; STREET: 6300 Cumbria Center, 701 Fifth Ave.
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,010
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

```
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.440C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-073-010-143

Query Match 100.0%; Score 468; DB 4; Length 97;
Best Local Similarity 100.0%; Pred. No. 8.7e-53;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPQLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAAFQAAHARFVA 60
DB 1 MSLLDAHIPQLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAAFQAAHARFVA 60
QY 61 AAAXVNTLLDVAQANLGEAAAGTYVAADAAAASYTGF 97
DB 61 AAAXVNTLLDVAQANLGEAAAGTYVAADAAAASYTGF 97

RESULT 5
US-09-287-849-19
; Sequence 19, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-19

Query Match 100.0%; Score 468; DB 4; Length 299;
Best Local Similarity 100.0%; Pred. No. 3.8e-52;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPQLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAAFQAAHARFVA 60
DB 189 MSLLDAHIPQLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAAFQAAHARFVA 248
QY 61 AAAXVNTLLDVAQANLGEAAAGTYVAADAAAASYTGF 97

; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.440C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 143:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 97 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-073-010-143

Query Match 100.0%; Score 468; DB 4; Length 97;
Best Local Similarity 100.0%; Pred. No. 8.7e-53;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPQLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAAFQAAHARFVA 60
DB 1 MSLLDAHIPQLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAAFQAAHARFVA 60
QY 61 AAAXVNTLLDVAQANLGEAAAGTYVAADAAAASYTGF 97
DB 61 AAAXVNTLLDVAQANLGEAAAGTYVAADAAAASYTGF 97

RESULT 6
US-09-287-849-14
; Sequence 14, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-14

Query Match 100.0%; Score 468; DB 4; Length 433;
Best Local Similarity 100.0%; Pred. No. 6.2e-52;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPQLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAAFQAAHARFVA 60
DB 335 MSLLDAHIPQLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESSAAAFQAAHARFVA 394
QY 61 AAAXVNTLLDVAQANLGEAAAGTYVAADAAAASYTGF 97
DB 395 AAAXVNTLLDVAQANLGEAAAGTYVAADAAAASYTGF 431

RESULT 7
US-09-287-849-16
; Sequence 16, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-16
```

; PRIOR APPLICATION NUMBER: US 09/025,197
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 09/056,556
 ; PRIOR FILING DATE: 1998-04-07
 ; PRIOR APPLICATION NUMBER: US 09/223,040
 ; PRIOR FILING DATE: 1998-12-30
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: Patent in Ver. 2.1
 ; SEQ ID NO 16
 ; LENGTH: 710
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
 US-09-287-849-16

 Query Match 100.0%; Score 468; DB 4; Length 710;
 Best Local Similarity 100.0%; Pred. No. 1.2e-51;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 Qy 1 MSLDDAHIPOLVASQSAFAAKAGLMRHTTIGAEQAAMSAQAAPHQGSAAFAQAAHARFVA 60
 Db 189 MSLDDAHIPOLVASQSAFAAKAGLMRHTTIGAEQAAMSAQAAPHQGSAAFAQAAHARFVA 248

 Qy 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97
 Db 249 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 285

 RESULT 8
 US-09-287-849-12
 ; Sequence 12, Application US/09287849
 ; Patent No. 6627198
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
 ; TITLE OF INVENTION: and Their Uses
 ; FILE REFERENCE: 014058-009020US
 ; CURRENT APPLICATION NUMBER: US/09/287,849
 ; CURRENT FILING DATE: 1999-04-07
 ; PRIOR APPLICATION NUMBER: US 08/818,112
 ; PRIOR FILING DATE: 1997-03-13
 ; PRIOR APPLICATION NUMBER: US 08/942,578
 ; PRIOR FILING DATE: 1997-10-01
 ; PRIOR APPLICATION NUMBER: US 09/025,197
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 09/056,556
 ; PRIOR FILING DATE: 1998-04-07
 ; PRIOR APPLICATION NUMBER: US 09/223,040
 ; PRIOR FILING DATE: 1998-12-30
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: Patent in Ver. 2.1
 ; SEQ ID NO 12
 ; LENGTH: 856
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:penta-fusion
 US-09-287-849-12

QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97
 Db 395 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 431

 RESULT 9
 US-08-818-112-88
 ; Sequence 88, Application US/08818112
 ; Patent No. 6290969
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 153
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA: US/08/818,112
 ; FILING DATE: 13-MAR-1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.411C6
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 88:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 95 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 US-08-818-112-88

 Query Match 18.7%; Score 87.5; DB 3; Length 95;
 Best Local Similarity 34.9%; Pred. No. 0.00071;
 Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;

 QY 18 FAKAGLMRHTTIGAEQAAMSAQAAPHQGSAAFAQAAHARFVAQAAKNTLLDVAQANL 77
 Db 13 FERISGDLTKIQDQVESTAGSLQGWGGAAGTAQAQAAVVRFOEAANKQKQELDEISTNIR 72

 QY 78 EAAGTYVAAD---AAAATYTG 97
 Db 73 QAGVOYSGRADEEQOQALSSQMGF 95

 RESULT 10
 US-08-818-111-89
 ; Sequence 89, Application US/08818111
 ; Patent No. 6338852
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonio

```
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-111-89
;
; Query Match 18.7%; Score 87.5; DB 4; Length 95;
; Best Local Similarity 34.9%; Pred.No.0.00071;
; Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;
;
; QY 18 FAKAGLMRHTIGCAEQAAWSAQAFHQGESSAFAQAAHAFVAAAKYNTLLDVAQANLG 77
; Db 13 FERISGLTKTQIDQVESTAGSLQGWGAAGTAAQAAVVRFOEAANKQKQELDEISTNIR 72
;
; QY 78 EAAGTYVAAD---AAAASYTYGF 97
; Db 73 QAGVQYSRADEEQQALSSQMGF 95
;
; RESULT 11
; US-09-056-556-88
; Sequence 88, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
;
; QY 18 FAKAGLMRHTIGCAEQAAWSAQAFHQGESSAFAQAAHAFVAAAKYNTLLDVAQANLG 77
; Db 13 FERISGLTKTQIDQVESTAGSLQGWGAAGTAAQAAVVRFOEAANKQKQELDEISTNIR 72
;
; QY 78 EAAGTYVAAD---AAAASYTYGF 97
; Db 73 QAGVQYSRADEEQQALSSQMGF 95
;
; RESULT 12
; US-09-072-596-89
; Sequence 89, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
; US-09-056-556-88
;
; Query Match 18.7%; Score 87.5; DB 4; Length 95;
; Best Local Similarity 34.9%; Pred.No.0.00071;
; Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;
;
; QY 18 FAKAGLMRHTIGCAEQAAWSAQAFHQGESSAFAQAAHAFVAAAKYNTLLDVAQANLG 77
; Db 13 FERISGLTKTQIDQVESTAGSLQGWGAAGTAAQAAVVRFOEAANKQKQELDEISTNIR 72
;
; QY 78 EAAGTYVAAD---AAAASYTYGF 97
; Db 73 QAGVQYSRADEEQQALSSQMGF 95
;
; RESULT 12
; US-09-072-596-89
; Sequence 89, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: amino acid
```

STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-596-89

Query Match 18.7%; Score 87.5; DB 4; Length 95;
Best Local Similarity 34.9%; Pred. No. 0.00071;
Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;

QY 18 FAKAGLMRHTIGQAEQAAMSAQAFHOGESSAAFAQAAHARFVAAAKVNTLLDVAQANLG 77

Db 13 FERISGDLKTIQIDVESTAGSLQGWGAAGTAAQAAVVRFQEAANKQKQELDEISTNIR 72

QY 78 EAAGTYVAAD---AAASTYTG 97

Db 73 QAGVQYSRADDEEQQALSSQMGF 95

RESULT 13

US-09-072-967-88

; Sequence 88, Application US/09072967

; Patent No. 6592877

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Houghton, Raymond

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Twardzik, Daniel R.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Hendrickson, Ronald C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

; AND DIAGNOSIS OF TUBERCULOSIS

; NUMBER OF SEQUENCES: 355

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/072,967

; FILING DATE: 05-MAY-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.411C9

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 88:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 95 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-09-072-967-88

Query Match 18.7%; Score 87.5; DB 4; Length 95;

Best Local Similarity 34.9%; Pred. No. 0.00071;

Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;

QY 18 FAKAGLMRHTIGQAEQAAMSAQAFHOGESSAAFAQAAHARFVAAAKVNTLLDVAQANLG 77

Db 13 FERISGDLKTIQIDVESTAGSLQGWGAAGTAAQAAVVRFQEAANKQKQELDEISTNIR 72

QY 78 EAAGTYVAAD---AAASTYTG 97

Db 73 QAGVQYSRADDEEQQALSSQMGF 95

RESULT 14

US-08-818-112-115

; Sequence 115, Application US/08818112

; Patent No. 6290369

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Houghton, Raymond

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Twardzik, Daniel R.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

; AND DIAGNOSIS OF TUBERCULOSIS

; NUMBER OF SEQUENCES: 153

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/818,112

; FILING DATE: 13-MAR-1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.411C6

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 115:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 100 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

US-08-818-112-115

Query Match 18.7%; Score 87.5; DB 3; Length 100;

Best Local Similarity 34.9%; Pred. No. 0.00076;

Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;

QY 18 FAKAGLMRHTIGQAEQAAMSAQAFHOGESSAAFAQAAHARFVAAAKVNTLLDVAQANLG 77

Db 18 FERISGDLKTIQIDVESTAGSLQGWGAAGTAAQAAVVRFQEAANKQKQELDEISTNIR 77

QY 78 EAAGTYVAAD---AAASTYTG 97

Db 78 QAGVQYSRADDEEQQALSSQMGF 100

RESULT 15

US-08-818-111-110

; Sequence 110, Application US/08818111

; Patent No. 6338852

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Campos-Neto, Antonio

APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 110:
SEQUENCE CHARACTERISTICS:
LENGTH: 100 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-818-111-110

Query Match 18.7%; Score 87.5; DB 4; Length 100;
Best Local Similarity 34.9%; Pred. No. 0.00076;
Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;
QY 18 FAKAGLMRHTTCQAEQAANSAQAPHQGESSAAFOAAHARFVAAAKYNTLLDVAQANLG 77
Db 18 FERISGLTKTQIDQVESTAGSLQGGWRGAAGTAQAQAAVVRFOEAANKOKQELDEISTNIR 77
QY 78 EAAGTYVAAD--AAAATYTG 97
Db 78 QAGVQYSRADEEQQALSSQMGF 100

Search completed: June 23, 2004, 17:10:43
Job time : 38 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2004, 17:12:36 ; Search time 43 Seconds
(without alignments)
636.847 Million cell updates/sec

Title: US-09-886-349A-24

Perfect score: 468
Sequence: 1 MSLDHPQLVQSAPFAA.....EAGTYVAADAAASTVTGF 97

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

Total number of hits satisfying chosen parameters: 1163542

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	468	100.0	97	US-09-073-009-109	Sequence 109, Appl
2	468	100.0	97	US-09-073-009-143	Sequence 143, Appl
3	468	100.0	97	US-09-793-306-109	Sequence 109, Appl
4	468	100.0	97	US-09-793-306-143	Sequence 143, Appl
5	468	100.0	97	US-09-886-349A-24	Sequence 24, Appl
6	468	100.0	97	US-10-098-732A-24	Sequence 24, Appl
7	468	100.0	299	US-09-287-849-19	Sequence 19, Appl
8	468	100.0	299	US-09-886-349A-47	Sequence 47, Appl
9	468	100.0	299	US-10-359-460-19	Sequence 19, Appl
10	468	100.0	299	US-10-098-732A-47	Sequence 47, Appl
11	468	100.0	433	US-09-287-849-14	Sequence 14, Appl
12	468	100.0	433	US-10-359-460-14	Sequence 14, Appl
13	468	100.0	710	US-09-287-849-16	Sequence 16, Appl
14	468	100.0	710	US-09-886-349A-49	Sequence 49, Appl
15	468	100.0	710	US-10-359-460-16	Sequence 16, Appl

Sequence 49, Appl
Sequence 12, Appl
Sequence 12, Appl
Sequence 300, Appl
Sequence 37, Appl
Sequence 37, Appl
Sequence 6207, A
Sequence 61942, A
Sequence 35, Appl
Sequence 89, Appl
Sequence 88, Appl
Sequence 35, Appl
Sequence 5, Appl
Sequence 639, Appl
Sequence 110, Appl
Sequence 115, Appl
Sequence 8, Appl
Sequence 10, Appl
Sequence 209, Appl
Sequence 346, Appl
Sequence 314, Appl
Sequence 351, Appl
Sequence 10, Appl
Sequence 112, Appl
Sequence 117, Appl
Sequence 124, Appl
Sequence 124, Appl
Sequence 64750, A
Sequence 26, Appl

710 14 US-10-098-732A-49
856 9 US-09-287-849-12
856 14 US-10-359-460-12
856 14 US-10-080-170-300
352 75.2 98 14 US-10-393-449-37
104 12 US-10-177-725-37
104 12 US-10-282-122A-62027
134 14 US-10-282-122A-61942
140 12 US-09-886-349A-35
145 12 US-10-193-002-89
145 14 US-10-084-843-88
145 14 US-10-098-732A-35
145 14 US-10-140-045-5
100 14 US-10-080-170-639
100 14 US-10-193-002-115
100 14 US-10-084-843-115
358 9 US-09-287-849-8
358 14 US-10-359-460-8
802 9 US-09-287-849-10
802 14 US-10-193-002-209
802 14 US-10-193-002-346
802 14 US-10-084-843-214
802 14 US-10-084-843-351
802 14 US-10-359-460-10
80 14 US-10-193-002-112
80 14 US-10-084-843-117
18 9 US-09-073-009-124
18 9 US-09-793-306-124
461 12 US-10-282-122A-64750
104 12 US-10-393-449-26

ALIGNMENTS

RESULT 1

US-09-073-009-109

; Sequence 109, Application US/09073009

; Patent No. US20010012888A1

; GENERAL INFORMATION:

; APPLICANT: Alderson, Mark

; APPLICANT: Dillon, Davin C.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Campos-Neto, Antonio

; TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF

; NUMBER OF SEQUENCES: 144

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY

; STREET: 6300 Columbia Center, 701 Fifth Ave.

; CITY: Seattle

; STATE: Washington

; COUNTRY: US

; ZIP: 98104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/073.009

; FILING DATE: 05-MAY-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.441C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 206-622-4900

; TELEFAX: 206-682-6031

; INFORMATION FOR SEQ ID NO: 109:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 97 amino acids

; TYPE: amino acid

TUBERCULOSIS AND MET

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-073-009-109

Query Match 100.0%; Score 468; DB 9; Length 97;
Best Local Similarity 100.0%; Pred. No. 4.4e-47;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAKAGLMRHTIGQAEOAAMSQAQFHQGSAAFOAAHARFVA 60

DB 1 MSLLDAHIPOLVASQSAFAKAGLMRHTIGQAEOAAMSQAQFHQGSAAFOAAHARFVA 60

QY 61 AAQVNTLLDVAQNLGEAGTGYVAADAAAATYTG 97

DB 61 AAQVNTLLDVAQNLGEAGTGYVAADAAAATYTG 97

RESULT 2

US-09-073-009-143

Sequence 143, Application US/09073009

Patent No. US2001001288A1

GENERAL INFORMATION:

APPLICANT: Alderson, Mark

APPLICANT: Dillon, David C.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Campos-Neto, Antonio

TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF

NUMBER OF SEQUENCES: 144

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY

STREET: 6300 Columbia Center, 701 Fifth Ave.

CITY: Seattle

STATE: Washington

COUNTRY: US

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/073,009

FILING DATE: 05-MAY-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.441C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-622-4900

TELEFAX: 206-682-6031

INFORMATION FOR SEQ ID NO: 143:

SEQUENCE CHARACTERISTICS:

LENGTH: 97 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-073-009-143

Query Match 100.0%; Score 468; DB 9; Length 97;
Best Local Similarity 100.0%; Pred. No. 4.4e-47;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAKAGLMRHTIGQAEOAAMSQAQFHQGSAAFOAAHARFVA 60

DB 1 MSLLDAHIPOLVASQSAFAKAGLMRHTIGQAEOAAMSQAQFHQGSAAFOAAHARFVA 60

QY 61 AAQVNTLLDVAQNLGEAGTGYVAADAAAATYTG 97

DB 61 AAQVNTLLDVAQNLGEAGTGYVAADAAAATYTG 97

TUBERCULOSIS AND ME

RESULT 3

US-09-793-306-109

Sequence 109, Application US/09793306

Patent No. US20020098200A1

GENERAL INFORMATION:

APPLICANT: Campos-Neto, Antonio

APPLICANT: Skeiky, Yasir

APPLICANT: Overdale, Pamela

APPLICANT: Jen, Shyian

APPLICANT: Lodes, Michael

APPLICANT: Corixa Corporation

TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy

TITLE OF INVENTION: of Tuberculosis

FILE REFERENCE: 014058-008740US

CURRENT APPLICATION NUMBER: US/09/793,306

CURRENT FILING DATE: 2001-02-26

PRIOR APPLICATION NUMBER: US 60/185,037

PRIOR FILING DATE: 2000-02-25

PRIOR APPLICATION NUMBER: US 60/223,828

PRIOR FILING DATE: 2000-08-08

NUMBER OF SEQ ID NOS: 164

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 109

LENGTH: 97

TYPE: PRT

ORGANISM: Mycobacterium tuberculosis

FEATURE:

OTHER INFORMATION: Tb472 (MSL)

US-09-793-306-109

Query Match 100.0%; Score 468; DB 9; Length 97;

Best Local Similarity 100.0%; Pred. No. 4.4e-47;

Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAKAGLMRHTIGQAEOAAMSQAQFHQGSAAFOAAHARFVA 60

DB 1 MSLLDAHIPOLVASQSAFAKAGLMRHTIGQAEOAAMSQAQFHQGSAAFOAAHARFVA 60

QY 61 AAQVNTLLDVAQNLGEAGTGYVAADAAAATYTG 97

DB 61 AAQVNTLLDVAQNLGEAGTGYVAADAAAATYTG 97

RESULT 4

US-09-793-306-143

Sequence 143, Application US/09793306

Patent No. US20020098200A1

GENERAL INFORMATION:

APPLICANT: Campos-Neto, Antonio

APPLICANT: Skeiky, Yasir

APPLICANT: Overdale, Pamela

APPLICANT: Jen, Shyian

APPLICANT: Lodes, Michael

APPLICANT: Corixa Corporation

TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy

TITLE OF INVENTION: of Tuberculosis

FILE REFERENCE: 014058-008740US

CURRENT APPLICATION NUMBER: US/09/793,306

CURRENT FILING DATE: 2001-02-26

PRIOR APPLICATION NUMBER: US 60/185,037

PRIOR FILING DATE: 2000-02-25

PRIOR APPLICATION NUMBER: US 60/223,828

PRIOR FILING DATE: 2000-08-08

NUMBER OF SEQ ID NOS: 164

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 143

LENGTH: 97

TYPE: PRT

ORGANISM: Mycobacterium tuberculosis

FEATURE:

OTHER INFORMATION: Mtb9.8

US-09-793-306-143

```
;
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB9.8 (MSL)
US-10-098-732A-24

Query Match      100.0%; Score 468; DB 9; Length 97;
Best Local Similarity 100.0%; Pred. No. 4.4e-47;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAAFQAAHARFVA 60
DB 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAAFQAAHARFVA 60

QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGTG 97
DB 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGTG 97

RESULT 5
US-09-886-349A-24
; Sequence 24, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB9.8 (MSL)
US-09-886-349A-24

Query Match      100.0%; Score 468; DB 12; Length 97;
Best Local Similarity 100.0%; Pred. No. 4.4e-47;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAAFQAAHARFVA 60
DB 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAAFQAAHARFVA 60

QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGTG 97
DB 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGTG 97

RESULT 6
US-10-098-732A-24
; Sequence 24, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
```

```
;
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB9.8 (MSL)
US-10-098-732A-24

Query Match      100.0%; Score 468; DB 14; Length 97;
Best Local Similarity 100.0%; Pred. No. 4.4e-47;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAAFQAAHARFVA 60
DB 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAAFQAAHARFVA 60

QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGTG 97
DB 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGTG 97

RESULT 7
US-09-287-849-19
; Sequence 19, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillion, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-19

Query Match      100.0%; Score 468; DB 9; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.7e-46;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAAFQAAHARFVA 60
DB 189 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAAFQAAHARFVA 248

QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGTG 97
DB 249 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGTG 285

RESULT 8
US-09-886-349A-47
; Sequence 47, Application US/09886349A
; Publication No. US20040086523A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 47
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
; OTHER INFORMATION: protein DPV-MTI-MSL (designated MTB31F) CDNA
US-09-886-349A-47

```

```

Query Match      100.0%; Score 468; DB 12; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.7e-46;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTICQAEQAAMSAQAQFHQGESAAFAQAAHARFVA 60
    |||||
DB 189 MSLLDAHIPOLVASQSAFAAKAGLMRHTICQAEQAAMSAQAQFHQGESAAFAQAAHARFVA 248

QY 61 AAAXVNTLLDVAQNGLGEAGTYVAADAAAATYTG 97
    |||||
DB 249 AAAXVNTLLDVAQNGLGEAGTYVAADAAAATYTG 285

```

```

RESULT 9
US-10-359-460-19
; Sequence 19, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR FILING DATE: 2003-02-05
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion

```

```

US-10-359-460-19
Query Match      100.0%; Score 468; DB 14; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.7e-46;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTICQAEQAAMSAQAQFHQGESAAFAQAAHARFVA 60
    |||||
DB 189 MSLLDAHIPOLVASQSAFAAKAGLMRHTICQAEQAAMSAQAQFHQGESAAFAQAAHARFVA 248

QY 61 AAAXVNTLLDVAQNGLGEAGTYVAADAAAATYTG 97
    |||||
DB 249 AAAXVNTLLDVAQNGLGEAGTYVAADAAAATYTG 285

```

```

RESULT 10
US-10-098-732A-47
; Sequence 47, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 47
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
; OTHER INFORMATION: protein DPV-MTI-MSL (designated MTB31F)
US-10-098-732A-47

```

```

Query Match      100.0%; Score 468; DB 14; Length 299;
Best Local Similarity 100.0%; Pred. No. 1.7e-46;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTICQAEQAAMSAQAQFHQGESAAFAQAAHARFVA 60
    |||||
DB 189 MSLLDAHIPOLVASQSAFAAKAGLMRHTICQAEQAAMSAQAQFHQGESAAFAQAAHARFVA 248

QY 61 AAAXVNTLLDVAQNGLGEAGTYVAADAAAATYTG 97
    |||||
DB 249 AAAXVNTLLDVAQNGLGEAGTYVAADAAAATYTG 285

```

```

RESULT 11
US-09-287-849-14
; Sequence 14, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13

```

```

; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-14

Query Match      100.0%; Score 468; DB 9; Length 433;
Best Local Similarity 100.0%; Pred. No. 2.7e-46;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSLLDAHIPQLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQAFHQGESSAAFOAAHARFVA 60
Db 335 MSLLDAHIPQLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQAFHQGESSAAFOAAHARFVA 394

Qy 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGF 97
Db 395 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGF 431

RESULT 12
US-10-359-460-14
; Sequence 14, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; PRIOR FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-10-359-460-14

Query Match      100.0%; Score 468; DB 14; Length 433;
Best Local Similarity 100.0%; Pred. No. 2.7e-46;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 MSLLDAHIPQLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQAFHQGESSAAFOAAHARFVA 60
Db 335 MSLLDAHIPQLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQAFHQGESSAAFOAAHARFVA 394

Qy 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGF 97
Db 395 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGF 431

RESULT 13
US-09-287-849-16
; Sequence 16, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 16
; LENGTH: 710
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
US-09-287-849-16

Query Match      100.0%; Score 468; DB 9; Length 710;
Best Local Similarity 100.0%; Pred. No. 4.9e-46;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSLLDAHIPQLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQAFHQGESSAAFOAAHARFVA 60
Db 189 MSLLDAHIPQLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQAFHQGESSAAFOAAHARFVA 248

Qy 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGF 97
Db 249 AAANKVNTLLDVAQANLGEAAGTYVAADAAAASYTGF 285

RESULT 14
US-09-886-349A-49
; Sequence 49, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-05-20

```

; PRIOR APPLICATION NUMBER: US 60/265,737
 ; PRIOR FILING DATE: 2001-02-01
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 49
 ; LENGTH: 710
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
 ; OTHER INFORMATION: protein DPV-MTI-MSL-WTCC#2 (designated MTb71F)
 US-09-886-349A-49

Query Match 100.0%; Score 468; DB 12; Length 710;
 Best Local Similarity 100.0%; Pred. No. 4.9e-46;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSLLDAHIPQIVASQSAFAKAGLMRHTIGQAEQAAMSAQAFHQGESSAAAFQAAHARFVA 60
 Db 189 MSLLDAHIPQIVASQSAFAKAGLMRHTIGQAEQAAMSAQAFHQGESSAAAFQAAHARFVA 248
 QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97
 Db 249 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 285

RESULT 15
 US-10-359-460-16
 ; Sequence 16, Application US/10359460
 ; Publication No. US20030147911A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Aldersen, Mark
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
 ; FILE REFERENCE: 014058-009020US
 ; CURRENT APPLICATION NUMBER: US/10/359,460
 ; CURRENT FILING DATE: 2003-02-05
 ; PRIOR APPLICATION NUMBER: US/09/287,849
 ; PRIOR FILING DATE: 1999-04-07
 ; PRIOR APPLICATION NUMBER: US 08/818,112
 ; PRIOR FILING DATE: 1997-03-13
 ; PRIOR APPLICATION NUMBER: US 08/942,578
 ; PRIOR FILING DATE: 1997-10-01
 ; PRIOR APPLICATION NUMBER: US 09/025,197
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 09/056,556
 ; PRIOR FILING DATE: 1998-04-07
 ; PRIOR APPLICATION NUMBER: US 09/223,040
 ; PRIOR FILING DATE: 1998-12-30
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 16
 ; LENGTH: 710
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:tetra-fusion
 US-10-359-460-16

Query Match 100.0%; Score 468; DB 14; Length 710;
 Best Local Similarity 100.0%; Pred. No. 4.9e-46;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSLLDAHIPQIVASQSAFAKAGLMRHTIGQAEQAAMSAQAFHQGESSAAAFQAAHARFVA 60
 Db 189 MSLLDAHIPQIVASQSAFAKAGLMRHTIGQAEQAAMSAQAFHQGESSAAAFQAAHARFVA 248
 QY 61 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 97

Db 249 AAANKVNTLLDVAQANLGEAAGTYVAADAAAATYTG 285
 Search completed: June 23, 2004, 17:18:23
 Job time : 48 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2004, 17:09:52 ; Search time 21 Seconds

(without alignments)
444.313 Million cell updates/sec

Title: US-09-886-349A-24

Perfect score: 468
Sequence: 1 MSLLDAHIPQLVASQSAFAA.....BAAGTYVAADAAASTYTGTF 97

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: Pir1:*
2: Pir2:*
3: Pir3:*
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	468	100.0	97	E70836	probable transcript
2	442	94.4	97	G70857	probable PE protein
3	356	76.1	98	A87226	PE-family protein
4	92	19.7	135	A70659	hypothetical prote
5	87.5	18.7	100	H70802	hypothetical glyci
6	87	18.6	1156	T34852	probable secreted
7	86	18.4	331	F70820	hypothetical glyci
8	86	18.4	431	F70571	hypothetical glyci
9	85	18.2	1147	T35781	hypothetical prote
10	84	17.9	457	H70820	hypothetical glyci
11	80.5	17.2	361	G70682	hypothetical glyci
12	80.5	17.2	394	G70881	probable PPE prote
13	80	17.1	405	F82181	galactokinase VC15
14	77.5	16.6	333	F70329	probable PPE prote
15	77.5	16.6	334	A70504	probable PPE prote
16	77.5	16.6	1238	T03465	probable exonuclea
17	76.5	16.3	175	A70392	hypothetical glyci
18	76	16.2	376	AG0392	tolA protein limpo
19	76	16.2	891	JN0867	peroxinectin-like
20	76	16.2	894	PN0567	peroxinectin-like
21	75.5	16.1	205	S19114	cgr-1 protein - C
22	75.5	16.1	582	IKESCA	colicin A - Citrob
23	75.5	16.1	882	G70812	hypothetical glyci
24	75	16.0	99	G70560	probable PE protei
25	75	16.0	339	JT0756	group-V allergen 1
26	75	16.0	1137	G70868	probable regulator
27	74.5	15.9	436	G81655	conserved hypothet
28	74.5	15.9	498	G70720	hypothetical glyci
29	74.5	15.9	582	B70953	hypothetical glyci

30	74.5	15.9	741	2	G70317	hypothetical glyci
31	74.5	15.9	1538	2	H70846	hypothetical glyci
32	74	15.8	97	2	S02376	antifreeze protein
33	74	15.8	401	1	QXBPII	hypothetical prote
34	74	15.8	978	2	C59237	pyruvate dehydroge
35	73.5	15.7	102	2	C70836	probable PE protei
36	73.5	15.7	558	2	E70756	hypothetical glyci
37	73.5	15.7	646	2	T42396	hypothetical prote
38	73.5	15.7	810	2	B84185	cytochrome-like pr
39	73.5	15.7	2082	2	T37056	probable multi-dom
40	73	15.6	357	2	G70773	probable prfa prot
41	73	15.5	805	2	T25795	hypothetical prote
42	72.5	15.5	801	2	F70824	hypothetical glyci
43	72.5	15.5	837	2	E70835	hypothetical glyci
44	72.5	15.5	957	2	D70835	hypothetical glyci
45	72.5	15.5	987	2	E70808	probable PPE prote

ALIGNMENTS

RESULT 1

E70836
probable transcription regulator Rv0287 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: E70836
R:Colt, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70836
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-97 <COL>
A:Cross-references: GB:AL021930; GB:AL123456; NID:G3261524; PIDN:CAA17362.1; PID:el25247
A:Experimental source: strain H37RV
C:Genetics:
C:Gene: Rv0287

Query Match 100.0%; Score 468; DB 2; Length 97;

Best Local Similarity 100.0%; Pred. No. 1.3e-38;

Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLLDAHIPQLVASQSAPAAKAGLMRHTIGQAEQAAMSAQAAPHQGESAAAFQAARFVA 60

DB 1 MSLLDAHIPQLVASQSAPAAKAGLMRHTIGQAEQAAMSAQAAPHQGESAAAFQAARFVA 60

QY 61 AAANKVTLIDVAQNLGEAAGTYVAADAAASTYTGTF 97

DB 61 AAANKVTLIDVAQNLGEAAGTYVAADAAASTYTGTF 97

RESULT 2

G70857
probable PE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

R:Colt, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: G70857

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-97 <COL>

A;Cross-references: GB:AL021287; GB:AL123456; NID:g3261508; PIDN:CAA16105.1; PID:e123778
 A;Experimental source: strain H37Rv
 C;Genetics:
 A;Gene: PE

Query Match 94.4%; Score 442; DB 2; Length 97;
 Best Local Similarity 91.8%; Pred. No. 4.3e-36;
 Matches 89; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSLLDAHLPOLVASQSFAAKAGLMRHTIGQAQAAMSQAQFHQGSAAFAQAHARFVA 60
 DB 1 MSLLDAHLPOLVIAHTAFKAGLMRHTIGQAQAAMSQAQFHQGSAAFAQAHARFVA 60

QY 61 AAAXVNTLLDVAQANLGEAAAGTYVAADAAASTYTG 97
 DB 61 AAAXVNTLLDIAQANLGEAAAGTYVAADAAASTYTG 97

RESULT 3
 A87226
 PE-family protein [imported] - Mycobacterium leprae
 C;Species: Mycobacterium leprae
 C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C;Accession: A87226
 R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; H
 R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
 eam, M.A.; Rutherford, K.M.
 Nature 409, 1007-1011, 2001
 A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
 A;Title: Massive gene decay in the leprosy bacillus.
 A;Reference number: A86909; MUID:21126732; PMID:11234002
 A;Accession: A87226
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-98 <STO>
 A;Cross-references: GB:AL450380; NID:g13093794; PIDN:CAC32063.1; GSPDB:GN00147
 C;Genetics:
 A;Gene: ML2532

Query Match 76.1%; Score 356; DB 2; Length 98;
 Best Local Similarity 77.9%; Pred. No. 9.2e-28;
 Matches 74; Conservative 7; Mismatches 14; Indels 0; Gaps 0;

QY 1 MSLLDAHLPOLVASQSFAAKAGLMRHTIGQAQAAMSQAQFHQGSAAFAQAHARFVA 60
 DB 1 MSLLDVHLPOLVASESFAAKAALMRSQINQACEAISQAQFHQGSAAFAQAHARFVT 60

QY 61 AAAXVNTLLDVAQANLGEAAAGTYVAADAAASTYT 95
 DB 61 AAEXKINALDIAQHLGEAAETVATDATAASTYT 95

RESULT 4
 A70659
 hypothetical protein Rv2541 - Mycobacterium tuberculosis (strain H37Rv)
 C;Species: Mycobacterium tuberculosis
 C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C;Accession: A70659
 R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A;Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A;Reference number: A70500; MUID:98295987; PMID:9634230
 A;Accession: A70659
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-135 <COL>
 A;Cross-references: GB:283863; GB:AL123456; NID:g3261685; PIDN:CAB06197.1; PID:e290876;
 A;Experimental source: strain H37Rv
 C;Genetics:
 A;Gene: Rv2541

Query Match 19.7%; Score 92; DB 2; Length 135;
 Best Local Similarity 36.6%; Pred. No. 0.047;
 Matches 30; Conservative 8; Mismatches 28; Indels 16; Gaps 2;

QY 11 LVASQSFAAKAGLMRHTIGQAQAAMSQAQFHQGSAAFAQAHARFVAAAANKVNTLLD 70
 DB 59 LAGELSAVAAPSG:-----AVLSWQA-----NAVAVNAAHARAGAAAAAASARMR 102

QY 71 VAQANLGEAAAGTYVAADAAAS 92
 DB 103 ATAAALGQAARRYAGQDTAAAA 124

RESULT 5
 H70802
 hypothetical protein Rv3874 - Mycobacterium tuberculosis (strain H37Rv)
 C;Species: Mycobacterium tuberculosis
 C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C;Accession: H70802
 R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A;Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A;Reference number: A70500; MUID:98295987; PMID:9634230
 A;Accession: H70802
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-100 <COL>
 A;Cross-references: GB:AL022120; GB:AL123456; NID:g3261558; PIDN:CAA17966.1; PID:g2960226
 A;Experimental source: strain H37Rv
 C;Genetics:
 A;Gene: Rv3874

Query Match 18.7%; Score 87.5; DB 2; Length 100;
 Best Local Similarity 34.9%; Pred. No. 0.095;
 Matches 29; Conservative 8; Mismatches 43; Indels 3; Gaps 1;

QY 18 FAKAGLMRHTIGQAQAAMSQAQFHQGSAAFAQAHARFVAAAANKVNTLLDVAQANLG 77
 DB 18 PERISGLKTDIQVESTAGSLQGWGAAGTAQAQAAVVRFBAAKQKQELDEISTNIR 77

QY 78 EAAGTYVAAD---AAAASTYTG 97
 DB 78 QAGVYSGRADEEQQAALSSQMGF 100

RESULT 6
 T34852
 Probable secreted protein - Streptomyces coelicolor
 C;Species: Streptomyces coelicolor
 C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C;Accession: T34852
 R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, February 1999
 A;Reference number: 221559
 A;Accession: T34852
 A;Status: preliminary; translated from GB/EMBL/DBU
 A;Molecule type: DNA
 A;Residues: 1-1156 <OLI>
 A;Cross-references: EMBL:AL035478; PIDN:CAB36606.1; GSPDB:GN00070; SCOEDB:SC235.19
 A;Experimental source: strain A3(2)
 C;Genetics:
 A;Gene: SCOEDB:SC2G5.19

Query Match 19.6%; Score 87; DB 2; Length 1156;
 Best Local Similarity 37.6%; Pred. No. 1.3;
 Matches 35; Conservative 12; Mismatches 34; Indels 12; Gaps 4;

QY 11 LVASQSA--FAAKAGL-----MRHTIGQAQAAMSQAQFHQGSAAAF-QAAHARFVAAA 62

395 LAAAQATATCAAAAAGISAAATARDUSAAAAQQAIVAAQAAGAAQSEAAVAAAAAADAQA 454

Qy 63 AK-----VNTLDDVAQNLGEAAGTYVVAADAAA 91
| : : : : :
455 ARATKAANRAQSLANTAASAAAAARKAADSAAA 487

Db

RESULT 7

F70820

hypothetical glycine-rich protein Rv0978c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: F70820

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Sqaers, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:9825987; PMID:9634230

A:Accession: F70820

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-331 <COL>

A:Cross-references: GB:AL021999; GB:AL123456; NID:G3261538; PIDN:CAA17577.1; PID:e12539d

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rv0978c

```

Query Match      18.4%; Score 86; DB 2; Length 331;
Best Local Similarity 33.0%; Pred. No. 0.45;
Matches 29; Conservative 14; Mismatches 41; Indels 4; Gaps 2;

Qy 9 POLVAGSFAKAGLMRHTTIGQAEQAAMSAQFHQGESAAFQA---AHAF-FVAAA 64
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8 POLVSTAADAABARIGSAINTANTAAAAATTOVLAADDEVSTAALFGSHGCHYQAIS 67
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 65 VNTILDVAQNILGEAAAGTYVAADAAAAAS 92
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 68 VAAYQORFVLALSOAGSTYVAVEAAASAT 95
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 8
F70571
Hypothetical glycine-rich protein Rv2615c - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70571
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.
Nature 393, 537-544 1998
A:Authors, Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the Biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: F70571
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-461 <COL>
A:Cross-references: GB:Z95387; GB:AL123456; NID:G3261763; PIDN:CAB08627.1; PID:G2104298
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: Rv2615c
C:Superfamily: phosphate-glucose-rich cell wall peptidase 1

	Query Match	18.4%;	Score 86;	DB 2;	Length 461;
	Best Local Similarity	33.0%;	Pred. NO. 0.64;		
	Matches	29;	Conservative 14;	Mismatches 41;	Indels 4; Gaps 2;
Cy	9 POLVASQSAFAKAGLMRRTTIGAEQAAMSAQAFHQGESSAAFOA---AHAR-FVAAAAK	64			
Dd	8 POLVTAAADAARIQSANTANTAAATQTVALAADVEYSTAIALFGSHGHYOAISAO	67			

```

Qy      65  VNTLLDVAQANLGEAAAGTYTVADAAAS  92
      |         |         |         |
Db      68  VAAVQRFVLALSQAGSTVAVAEAAASAT  95

RESULT 9
T35781
hypothetical protein SC8A6.14c SC8A6.14c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35781
E:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1998
A:Reference number: Z21570
A:Accession: T35781
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1147 <SEE>
A:Cross-references: EMBL:AL031013; PIN:CAA19786.1; GSFDB:GN000070; SCOEDB:IS
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC8A6.14c

```

```

Query Match      18.2%; Score 85; DB 2; Length 1147;
Best Local Similarity 35.1%; Fred. No. 2;
Matches 34; Conservative 11; Mismatches 36; Indels 16; Gaps 3;

QY 12 VASQSFPAKAGLUMRTTIGQEQAAWS-----AQAFHQGESSAFAQAAHARFVAAAAK 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 312 VTAAIAAAAAAAAAAGRAASRAYSAATISAKNANMAEAAQAQAAKAYVMVDVTTAAKAK 371
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 65 VNTLL-DVAQNLGEAGTVV-----AADAASAS 92
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 372 LAALASDAEAEAGNAAAGAAVDASAAARASADAAAAAS 408
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 10

RESULT 10
H70820
hypotheical glycine-rich protein RV0980C - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:accession: H70820
R:Coler, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Soares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: H70820
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-457 <CO>
A:Cross-references: GB:AL021999; GB:AL123456; NID:g3261538; PIDN:CAA17579.1; PID:g2916941
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV0980C
C:Superfamily: Phaseolus glyicine-rich cell wall protein 1.8

```
Query Match      17.9%; Score 84; DB 2; Length 457;  
Best Local Similarity 31.8%; Pred.No. 0.99;  
Matches         Conservative 16; Mismatches 40; Indels    4; Gaps   2;  
  
QY       9 POLVASQSFAAKAGLMRHTIQAEQAAMSAQAFHQGESSAAFOA---AHAX-FVAAAAK 64  
          |||:::||||  
DB        8 POLVTAAAADAARIISAINANTANATAATTVQLAAAHDEVSTAIAALFGSHGHOYAISAQ 67  
          |::|:  
  
QY       65 VNTLLDVAQNIGEAAGTYVAADAAAAAS 92  
          ::|::|:  
db       68 VAAYGERFVLALSQASSITYAVAREEASAT 95  
          :|::|:
```

RESULT 11

G70682
 Hypothetical glycine-rich protein RV2396 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
 C:Accession: G70682
 R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: G70682
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-361 <COL>
 A: Cross-references: GB:Z81368; GB:AL123456; NID: g3261656; PIDN: CAB03730.1; PID: e279646;
 A: Experimental source: strain H37RV
 C: Genetics:
 A: Gene: RV2396

Query Match 17.2%; Score 80.5; DB 2; Length 361;
 Best Local Similarity 35.1%; Pred. No. 1.7; Mismatches 37; Indels 13; Gaps 3;
 Matches 33; Conservative 11

QY 11 LVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGE-----SSAFAQAHAAR-F 58
 Db 4 LIASPEALANTATYLT-CIGSAISANAANAAPTTEILTAGTDEVSTAISALFQAHAQAY 62

QY 59 VAAAKVNTLDDVAQNLGEAGTIVAAADAAAS 92
 Db 63 QALSARHAAAFHQDFVHTLTAGAGSYMAAERAAAS 96

RESULT 12
 G70881
 Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: G70881
 R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: G70881
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-394 <COL>
 A: Cross-references: GB:AL008967; GB:AL123456; NID: g3261491; PIDN: CAAL15564.1; PID: e117389
 A: Experimental source: strain H37RV
 C: Genetics:
 A: Gene: PPE

Query Match 17.2%; Score 80.5; DB 2; Length 394;
 Best Local Similarity 36.7%; Pred. No. 1.9; Mismatches 37; Indels 7; Gaps 5;
 Matches 33; Conservative 13

QY 9 POLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGE---SSAFAQAHAARFVAAAKV 65
 Db 107 PRAEVANRALLA-ALVATNVLCQNTPAIMATEA-HYGENWQAQDALMYGYAASSAAAGRL 164

QY 66 NTLDDVAQNL-GEAAGTIVAA-ADAAAST 93
 Db 165 NPLITPSQTANVAGLQAQAAGVSHAAAST 194

RESULT 13
 F82181
 galactokinase VC1595 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
 C:Species: Vibrio cholerae
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: A70504

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: F82181
 R: Heidelberg, J. F.; Eisen, J. A.; Nelson, W. C.; Clayton, R. A.; Gwinn, M. L.; Dodson, R. J.; Chardon, D.; Ermolaeva, M. D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P. I.; R. K.; Mekalanos, J. J.; Venter, J. C.; Fraser, C. M.
 Nature 406, 477-483, 2000
 A: Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A: Reference number: A82035; MUID: 20406833; PMID: 10952301
 A: Accession: F82181
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-405 <HEI>
 A: Cross-references: GB:AE004236; GB:AE003852; NID: g9656095; PIDN: AAP94749.1; GSPDB: GNC011
 A: Experimental source: serogroup O1; strain N16961; biotype El Tor
 C: Genetics:
 A: Gene: VC1595
 C: Superfamily: galactokinase

Query Match 17.1%; Score 80; DB 2; Length 405;
 Best Local Similarity 27.3%; Pred. No. 2.1; Mismatches 39; Indels 8; Gaps 3;
 Matches 24; Conservative 17

QY 2 SLDDAHIPOLVASQSAF-AAKAGLMRHTTIGQAEQAAMSAQAQFHQGES---SAFAQAHAAR 57
 Db 261 ALRDVTLAGLTAKQAELEDPVVAKBARHVITENRTLHAAQALREGNMPRLGELMAASHAS 320

QY 58 -----FVAAAANKVNTLDDVAQNLGEAAG 81
 Db 321 MRDDFEITVKEIDTLVEIVQSVIGDQGG 348

RESULT 14
 G70929
 Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: G70929
 R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A: Reference number: A70500; MUID: 98295987; PMID: 9634230
 A: Accession: G70929
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-393 <COL>
 A: Cross-references: GB:AL022021; GB:AL123456; NID: g3250699; PIDN: CAAL17711.1; PID: e125460
 A: Experimental source: strain H37RV
 C: Genetics:
 A: Gene: PPE

Query Match 16.8%; Score 77.5; DB 2; Length 393;
 Best Local Similarity 31.4%; Pred. No. 3.6; Mismatches 41; Indels 19; Gaps 4;
 Matches 32; Conservative 10

QY 9 POLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAQFHQGESAAFAQAHAARFV---AAAAKV 65
 Db 60 PASANMAEAVAPVVAWMSAAQAQAATQARA-----AAAFEAFAAATVPPPLIAANR 114

QY 66 NTLDDVAQNL-GEAAGTIVAA-----DAAASTYTG 96
 Db 115 ASLQMLISTNVFGQNTSAIAAAEAQYGENWQAQDAAMAYAYAG 156

RESULT 15
 A70504
 Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: A70504

Search completed: June 23, 2004, 17:13:38
Job time : 27 secs

OM protein - protein search, using sw model

Run on: June 23, 2004, 17:09:52 ; Search time 17 Seconds
(without alignments)
297.106 Million cell updates/sec

Title: US-09-886-349A-24
Perfect score: 468
Sequence: 1 MSLLDAHLPQLVASQAF...EAGTYVAADAAASTYTF 97

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	87.5	18.7	99	1 ESXB MYCTU
2	80	17.1	386	1 GALL_VIBCH
3	77.5	16.6	1238	1 SBCC_RHOCA
4	75.5	16.1	428	1 FXB2_MOUSE
5	75.5	16.1	592	1 CEA_CITFR
6	75	16.0	339	1 MP5B_LOLPR
7	74.5	15.9	222	1 DEOC_CORGL
8	74.5	15.9	436	1 Y868_CHLMU
9	74.5	15.9	498	1 Y118_MYCTU
10	74	15.8	97	1 ANP_LIMPE
11	74	15.8	774	1 STPL_LAMED
12	74	15.8	880	1 BRC4_DROME
13	73.5	15.7	558	1 YJ83_MYCTU
14	73	15.6	357	1 RFL1_MYCTU
15	72.5	15.5	668	1 PAU1_DROME
16	72.5	15.5	801	1 Y747_MYCTU
17	72.5	15.5	957	1 Y278_MYCTU
18	72.5	15.5	1023	1 R712_ACTPL
19	72	15.4	905	1 CTFP_MYCTU
20	71.5	15.3	263	1 PSF1_ARATH
21	71	15.2	324	1 OSA_DROSI
22	70.5	15.1	163	1 CU38_LOGMI
23	69.5	14.9	274	1 PYRF_MYCTU
24	69	14.7	276	1 PYRF_MYCTU
25	69	14.7	395	1 TRBL_AGRU
26	69	14.7	408	1 YS92_MYCTU
27	69	14.7	474	1 CYAE_BORPE
28	68.5	14.6	102	1 YD86_MYCTU
29	68.5	14.6	275	1 HUPH_RHOCA
30	68.5	14.6	2541	1 TLN1_HUMAN
31	68.5	14.6	2541	1 TLN1_MOUSE
32	68	14.5	866	1 AREA_ASPOR
33	67.5	14.4	678	1 YF48_MYCTU

34	67.5	14.4	1023	1 RT11_ACTPL
35	67	14.3	276	1 SX21_HUMAN
36	67	14.3	2716	1 OSA_DROME
37	66.5	14.2	355	1 YH11_ECOLI
38	66.5	14.2	360	1 ALF_DROME
39	66.5	14.2	415	1 FOX1_CAEEL
40	66.5	14.2	550	1 CCF_DROME
41	66.5	14.2	697	1 SIM_DROME
42	66.5	14.2	1120	1 STFR_ECOLI
43	66	14.1	176	1 VP26_EBV
44	66	14.1	441	1 ZRAR_ECO57
45	66	14.1	511	1 ZPM1_STRCO

ALIGNMENTS

RESULT 1

ESXB_MYCTU

ID ESXB_MYCTU STANDARD; PRT; 99 AA.

AC O69739;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ESAT-6 like protein esxB (10 kDa culture filtrate antigen cfp10)
DE (Secreted antigenic protein MTS-10)
GN ESXB OR CFP10 OR LHP OR RV3874 OR MT3988 OR MTV027.09 OR
GN MB3904.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=99061212; PubMed=9846755;
RA Berthet F.-X., Rasmussen P.B., Rosenkrands I., Andersen P.,
RA Gicquel B.;
RT "A Mycobacterium tuberculosis operon encoding ESAT-6 and a novel
RT low-molecular-mass culture filtrate protein (CFP-10).";
RL Microbiology 144:3195-3203 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fieischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis;
RA Singh B., Siddiqui Z., Singh S., Sharma P.;

```
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RT RL Nature 406:477-483(2000).
CC -!- CATALYTIC ACTIVITY: ATP + D-galactose = ADP + D-galactose 1-
CC phosphate
CC -!- PATHWAY: Galactose metabolism; first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Belongs to the GMPK kinase family. Galk subfamily.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outpost at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AE004236; AAF94749.1; -.
DR TIGR: VCL1595; -.
DR HAMAP: MF_00246; -.
DR InterPro: IPR000705; Galactokinase.
DR InterPro: IPR001174; Galkinase.
DR InterPro: IPR006204; GHMP Kinase.
DR InterPro: IPR006203; GHMPKase ATP.
DR InterPro: IPR006206; Mgv_galkinase.
DR Pfam: PF00288; GHMP_kinases; 1.
DR PRINTS: PR00473; GALCTOKINASE.
DR PRINTS: PRO0960; LMBPPROTEIN.
DR PRINTS: PRO0959; MEVGALKINASE.
DR TIGRFAMs: TIGR00131; gal kin; 1.
DR PROSITE: PS00106; GALACTOKINASE; 1.
DR PROSITE: PS00627; GHMP_KINASES_ATP; 1.
DR Transferrase; Kinase; Galactose metabolism; ATP-binding;
KW Complete proteome.
DR NP_BIND 123 133 ATP (POTENTIAL).
DR SEQUENCE 386 AA; 41933 MW; F1A77EB9F796DD CRC64;
SQ
Query Match 17.1%; Score 80; DB 1; Length 386;
Best Local Similarity 27.3%; Pred. No. 0.93;
Matches 24; Conservative 17; Mismatches 39; Indels 8; Gaps 3;
QY 2 SLLDAHIPQLVASOSAF-AAKAGLMRHITIGAEQAAMSAAQAFHGES---SAAFQAHAR 57
Ddb 242 ALRDVTLAQLTAKQALDLPVVKAKRVHTENRTLHAQAALREGNPRIGELMAASHAS 301
QY 58 ----FVAAAANKVTLLDVQAQNLGEAG 81
Ddb 302 MRDFEITVKEDITLVEIVQSVIDQGQ 329
```

```

RESULT 3
      SEQID      SBCC_RHOCA      STANDARD;      PRT;      1238 AA.
      O68032;
28-FEB-2003 (Rel. 41, Created)
28-FEB-2003 (Rel. 41, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Nuclease sbccD subunit C.
SBCC.
Rhodobacter capsulatus (Rhodospseudomonas capsulata).
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
Rhodobacteraceae; Rhodobacter.
NCBI_TaxID=1061;
[1]
SEQUENCE FROM N.A.
STRAIN=SB1003 / St Louis;
MEDLINE=97404404; PubMed=9256491;
Vitek C., Paces V., Maltsev N., Paces J., Haselkorn R., Fonstein M.;
"Sequence of a 189-kb segment of the chromosome of Rhodobacter
capsulatus SB1003";
Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).
-!- FUNCTION: SbcCD cleaves DNA hairpin structures. These structures
containing DNA

```

recombination reactions. The complex acts as a 3'->5', double strand exonuclease that can open hairpins. It also has a 5', single-strand endonuclease activity (By similarity).

-!- SUBUNIT: Heterodimer of sbcC and sbcD (By similarity).

-!- SIMILARITY: Belongs to the SMC family. SMC subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF010496; AAC16118.1; -

DR PIR; T03465; T03465.

DR InterPro; IPR003439; ABC transporter.

KW Hydrolyase; Nuclease; Exonuclease; Endonuclease; DNA replication; DNA recombination; ATP-binding; Coiled coil.

FT NP_BIND 37 44 ATP (POTENTIAL).

FT DOMAIN 395 438 COILED COIL (POTENTIAL).

FT DOMAIN 466 487 COILED COIL (POTENTIAL).

FT DOMAIN 521 600 COILED COIL (POTENTIAL).

FT DOMAIN 724 770 COILED COIL (POTENTIAL).

FT DOMAIN 901 943 COILED COIL (POTENTIAL).

FT DOMAIN 1019 1052 COILED COIL (POTENTIAL).

SQ SEQUENCE 1238 AA; 128046 MW; 256B6C63A859AE1 CRC64;

Query Match 16.6%; Score 77.5; DB 1; Length 1238;

Best Local Similarity 31.6%; Pred. No. 4.8;

Matches 24; Conservative 13; Mismatches 36; Indels 3; Gaps 1;

21 KAGLMRHITCOAQSAQMSAQFHCQSSAAFOAHARFVAAAKVNTLIDVAQANLGEAA 80

185 RATILERVCTGLYRAVSIRVYTERTEMA---RAEHAQLLARAHEHLLDDAFAALTEET 241

81 GTYVADAAASIVTG 96

242 AARTATAAATAEAG 257

RESULT 4

FXB2 MOUSE STANDARD; PRT; 428 AA.

AC Q64733;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Forkhead box protein B2 (transcription factor FKH-4).

CN FOXB2 OR FKH4.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97014266; PubMed=8861101;

RA Kaestner K.H., Schuetz G., Monaghan A.P.;

RT "Expression of the winged helix genes fkh-4 and fkh-5 defines domains in the central nervous system.";

RL Mech. Dev. 55:221-230(1996).

RN [2]

RP SEQUENCE OF 4-114 FROM N.A.

RC STRAIN=129;

RX MEDLINE=93361500; PubMed=7689224;

RA Kaestner K.H., Lee K.H., Schleindorff J., Hiemisch H., Monaghan A.P., Schuetz G.;

RT "Six members of the mouse forkhead gene family are developmentally regulated.";

RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- DEVELOPMENTAL STAGE: Expressed during embryogenesis.

CC -!- SIMILARITY: Contains 1 fork-head domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X92591; CAA63335.1; -

DR PIR; D47746; D47746.

DR HSSP; O63245; 24PH.

DR TRANSPAC; T02442;

DR MGD; MGI:1347468; Foxb2.

DR InterPro; IPR001766; TF_Fork_head.

DR Pfam; PF00250; Fork_head; 1.

DR PRINTS; PR00053; FORKHEAD.

DR ProDom; PD000425; TF_Fork_head; 1.

DR SMART; SM00339; FH; 1.

DR PROSITE; PS00657; FORK_HEAD_1; 1.

DR PROSITE; PS00658; FORK_HEAD_2; 1.

DR PROSITE; PS50039; FORK_HEAD_3; 1.

KW DNA-binding; Nuclear protein; Transcription regulation.

FT DNA_BIND 12 103 FORK-HEAD.

FT DOMAIN 139 153 POLY-HIS.

FT DOMAIN 156 162 POLY-HIS.

FT DOMAIN 163 172 POLY-PRO.

FT DOMAIN 217 231 POLY-ALA.

FT DOMAIN 249 258 POLY-ALA.

FT DOMAIN 321 330 POLY-ALA.

FT DOMAIN 336 399 POLY-ALA.

SQ SEQUENCE 428 AA; 45170 MW; DB8A8EFD1E94AB10 CRC64;

Query Match 16.1%; Score 75.5; DB 1; Length 428;

Best Local Similarity 37.1%; Pred. No. 2.8;

Matches 26; Conservative 6; Mismatches 31; Indels 7; Gaps 2;

31 QAEQAQMSAQAFHQG---ESSAAFOAHARFVAAAKVNTLIDVAQANLGEAGTYVAAD 87

198 QPQSQPPQTSHPGKMQEAAVAAAAAAGVSGRLSQFPFPGLSAA----AAA 253

88 AAAASTYTG 97

254 AAAAASSTG 263

RESULT 5

CEA_CITER STANDARD; PRT; 592 AA.

ID CEA_CITER

AC P04480;

DT 13-AUG-1987 (Rel. 05, Created)

DT 13-AUG-1987 (Rel. 05, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Colicin A.

GN CAA.

OS Citrobacter freundii.

OG Plasmid ColA-CR31.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Citrobacter.

OX NCBI_TaxID=546;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=84036205; PubMed=6313941;

RA Morlon J., Lloubes R., Varenne S., Chartier M., Lazdunski C.;

RT "Complete nucleotide sequence of the structural gene for colicin A, a gene translated at non-uniform rate.";

RL J. Mol. Biol. 170:271-285(1983).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=88174422; PubMed=2832701;

RA Morlon J., Chartier M., Bidaud M., Lazdunski C.;

RT "The complete nucleotide sequence of the colicinogenic plasmid ColA.

RT High extent of homology with Cole1.1;
 RL Mol. Gen. Genet. 211:231-243(1988).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 389-592.
 RX MEDLINE=92235820; PubMed=1373773;
 RA Parker M.W., Postma J.P.M., Patters F., Tucker A.D., Tsernoglou D.;
 RT "Refined structure of the pore-forming domain of colicin A at 2.4-A
 resolution.";
 RL J. Mol. Biol. 224:639-657(1992).
 CC -!- FUNCTION: This colicin is a channel-forming colicin. This class of
 transmembrane toxins depolarize the cytoplasmic membrane, leading
 to dissipation of cellular energy.
 CC -!- FUNCTION: COLICINS ARE POLYPEPTIDE TOXINS PRODUCED BY AND ACTIVE
 AGAINST ESCHERICHIA COLI AND CLOSELY RELATED BACTERIA.
 CC -!- SIMILARITY: Belongs to the channel forming colicin family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X01008; CAA25503.1; -;
 CC EMBL; M37402; AAA72879.1; -;
 CC PIR; I40784; IKEBCA.
 CC FDB; ICOL; 15-JUL-93.
 CC InterPro; IPR000293; Channel_colicin.
 CC Pfam; PF01024; Colicin; 1.
 CC PRINTS; PR00280; CHANCOLICIN
 CC ProDom; PR002657; Channel_colicin; 1.
 CC PROSITE; PS00276; CHANNEL_COLICIN; 1.
 KW Antibiotic; Bacteriocin; Plasmid; Transmembrane; 3D-structure.
 FT TRANSMEM 528 548
 FT TRANSMEM 555 575
 FT TRANSMEM 587 588
 FT HELIX 396 418
 FT HELIX 420 434
 FT TURN 435 435
 FT TURN 437 438
 FT HELIX 444 455
 FT TURN 456 456
 FT HELIX 458 460
 FT HELIX 464 475
 FT TURN 476 476
 FT HELIX 479 489
 FT TURN 491 493
 FT HELIX 496 497
 FT TURN 498 515
 FT HELIX 519 530
 FT TURN 531 532
 FT HELIX 535 552
 FT HELIX 557 575
 FT TURN 577 577
 FT HELIX 578 586
 FT TURN 587 588
 SQ SEQUENCE 592 AA; 62992 MW; B80FALF52A8CFCD CRC64;
 Query Match 16.1%; Score 75.5; DB 1; Length 592;
 Best Local Similarity 31.5%; Pred. No. 3.8; Indels 19; Gaps 4;
 Matches 28; Conservative 12; Mismatches 30;
 QY 8 IPOLVAS-----QSAFAAKGLMRHTTIGQAQAAMSAQAFAHQESSAAFAQAARF 58
 Db 174 VPQLTASVNEGMRTRQEAADRAAEANARALABEE-----ARAISGSKSAEPDAG-KRV 227
 QY 59 VAAAKTNTLLDVAQNLGEAAAGTYVAAD 87
 Db 228 ERAQAANT-----AQLNVLNLSGAVSAAN 252
 RESULT 6
 MP5B_LOLPR

ID MP5B_LOLPR STANDARD; PRT; 339 AA.
 AC Q40237;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Major pollen allergen Lol p 5b precursor (Lol p Vb).
 OS Lolium perenne (Perennial ryegrass).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Poaceae; Lolium.
 OC NCBI_TaxID=4522;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Pollen;
 RX MEDLINE=94085783; PubMed=8262382;
 RA Ong E.K., Griffith I.J., Knox R.B., Singh M.B.;
 RT "Cloning of a cDNA encoding a group-V (group-IX) allergen isoform
 from rye-grass pollen that demonstrates specific antigenic
 RT immunoreactivity.";
 RL Gene 134:235-240(1993).
 CC -!- SUBCELLULAR LOCATION: STARCH GRANULE (BY SIMILARITY).
 CC -!- TISSUE SPECIFICITY: Pollen, starch granules (BY similarity).
 CC -!- ALLERGEN: Causes an allergic reaction in human. Causes grass
 CC pollen allergy.
 CC -!- SIMILARITY: BELONGS TO THE POA P IX/PHL P VI FAMILY OF ALLERGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; LI3083; AAA333405.1; -;
 CC PIR; JT0756; JT0756.
 CC InterPro; IPR001778; POA_allergen_C.
 CC InterPro; IPR002914; POA_allergen_N.
 CC Pfam; PF01620; Pollen_allerg_2; 1.
 CC PRINTS; PR00833; POAALLERGEN.
 KW Signal; Allergen; Multigene family; Repeat.
 FT SIGNAL 1 25
 FT CHAIN 26 339 MAJOR POLLEN ALLERGEN LOL P 5B.
 FT DOMAIN 32 58 9 X 3 AA TANDEM REPEATS OF [PA]-A-[TA].
 FT REPEAT 32 34 1-1.
 FT REPEAT 35 37 1-2.
 FT REPEAT 38 40 1-3.
 FT REPEAT 41 43 1-4.
 FT REPEAT 44 46 1-5.
 FT REPEAT 47 49 1-6.
 FT REPEAT 50 52 1-7.
 FT REPEAT 53 55 1-8.
 FT REPEAT 56 58 1-9.
 FT DOMAIN 285 334 6 X 9 AA APPROXIMATE TANDEM REPEATS OF
 FT T-A-T-A-T-P-A-A-A.
 FT REPEAT 285 290 2-1 (INCOMPLETE).
 FT REPEAT 292 300 2-2.
 FT REPEAT 301 309 2-3.
 FT REPEAT 310 318 2-4.
 FT REPEAT 319 327 2-5.
 FT REPEAT 328 334 2-6 (INCOMPLETE).
 FT DOMAIN 81 87 POLY-ALA.
 FT DOMAIN 288 293 POLY-ALA.
 SQ SEQUENCE 339 AA; 33750 MW; 3CAA3115625B0610 CRC64;
 Query Match 16.0%; Score 75; DB 1; Length 339;
 Best Local Similarity 35.2%; Pred. No. 2.5; Indels 6; Gaps 4;
 Matches 32; Conservative 16; Mismatches 37;
 QY 8 IPOLVAS-QSAFAAK---AGLMRHTTIGQ-REQAAMSAQAFAHQESSAAFAQAARFVAAA 62
 Db 241 IPTLVAAVKQAYAAKQATAPVKYTVSETALKKAVTAMSEAKEATPA-AAATATPTPAA 299


```
QY 63 AKVNTLLDVAQANLGEAAGTYVAADAAAST 93
DE 16-OCT-2001 (Rel. 40, Last sequence update)
DB 300 ATATATPAAAYATATPAAATATATPAAATAT 330

RESULT 7
DECO_CORGL STANDARD; PRT; 222 AA.
AC Q8NTC4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Deoxyribose-phosphate aldolase (EC 4.1.2.4) (Phosphodeoxyriboaldolase)
DE (deoxyriboaldolase) (DERA).
DECO OR GLO383.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -I- CATALYTIC ACTIVITY: 2-deoxy-D-ribose 5-phosphate = D-
CC Glyceraldehyde 3-phosphate + acetaldehyde.
CC -I- PATHWAY: Nucleotide and deoxyribonucleotide catabolism.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: Belongs to the deoc/fbaB aldolase family. Deoc
CC subfamily 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP005275; BAB97776.1; -.
CC DR HAMAP; MF_00114; -.
CC DR InterPro; IPR002915; Deoc.
CC DR InterPro; IPR003009; FMN_enzyme.
CC DR Pfam; PF01791; Deoc; 1.
CC DR Lyase; Schifff base; Complete proteome.
CC FT BINDING 156 156 SCHIFF-BASE WITH ACETALDEHYDE (BY
CC SIMILARITY).
CC FT ACT_SITE 186 186 BY SIMILARITY.
CC SQ SEQUENCE 222 AA; 22520 MW; 563C3D6B6805DD08 CRC64;

Query Match 15.9%; Score 74.5; DB 1; Length 222;
Best Local Similarity 21.7%; Pred. No. 1.9;
Matches 28; Conservative 18; Mismatches 34; Indels 49; Gaps 3;

QY 8 IPQLVASQSFAKAGLMRHTIGQEQAAVSAQAFHGV--AAAKV 65
DE 16-OCT-2001 (Rel. 40, Last sequence update)
DB 300 ATATATPAAAYATATPAAATATATPAAATAT 330

RESULT 8
Y868_CHLMU STANDARD; PRT; 436 AA.
AC Q8FJG1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein TC0868.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / N199;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., Bass S.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Neilson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.W.;
RA Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -I- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0808/CT579/TC0868
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE002353; AAF39664.1; -.
CC DR PIR; G81655; G81655.
CC DR TIGR; TC0868; -.
CC KW Hypothetical protein; Complete proteome.
CC SQ SEQUENCE 436 AA; 43891 MW; 43521C9924CD7A48 CRC64;

Query Match 15.9%; Score 74.5; DB 1; Length 436;
Best Local Similarity 28.2%; Pred. No. 3.6;
Matches 33; Conservative 10; Mismatches 39; Indels 35; Gaps 3;

QY 10 QLVASQSFAKAGLMRHTIGQEQAAVSAQAFHQ----- 44
DB 163 QAIQNVAMAFEIGNAIRTQNHQAATLQAKQSLISGINTIVGFVAVSGGILSASKS 222
QY 45 --GESSAAF-----QAAHAFVAAAKVNTLLDVAQANLGEAAGTYVAADAAAST 93
DB 223 LGGLKSAFTNETATATASSATSAAKTAVNALDDV--ANVAATAGTKAASGAASAAS 277

RESULT 9
YI18_MYCTU STANDARD; PRT; 498 AA.
AC Q50615;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PE-PERS family protein RV1818C/MT1866.
GN RV1818C OR MT1866 OR MTCY1A11.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Biglieri K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellton S., Squares S., Squares R.,
```



```

RA Haggaard-Ljungquist E., Halling C., Calendar R.;
RT "DNA sequences of the tail fiber genes of bacteriophage P2: evidence
RT for horizontal transfer of tail fiber genes among unrelated
RT bacteriophages.";
RL J. Bacteriol. 174:1462-1477(1992).
RN [3]
RP RECONSTRUCTION.
RX MEDLINE=93068310; PubMed=1439923;
RA Hendrix R.W., Duda R.L.;
RT "Bacteriophage lambda PaPa: not the mother of all lambda phages.";
RL Science 258:1145-1148(1992).
CC -!- MISCELLANEOUS: The common laboratory strain of bacteriophage
CC lambda; lambda PaPa: carries a frameshift mutation relative to Ur-
CC lambda; the original isolate. The Ur-lambda virions have thin,
CC joined tail fibers (side tail fibers) that are absent from lambda
CC wild type. Relative to lambda PaPa, Ur-lambda has expanded
CC receptor specificity and adsorbs to E.coli cells more rapidly.
CC -!- SIMILARITY: Belongs to the tail fiber family.
CC -!- CAUTION: This is a conceptual translation; a frameshift was
CC corrected in position to 396 to recreate the original stf protein.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement.(See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; J02459; AAA96555.1; ALT FRAME.
CC ENBL; J02459; AAA96557.1; ALT FRAME.
CC PIR; C43009; QXBP2L.
CC InterPro; IPR008969; Carboxypeptid_reg.
CC InterPro; IPR005003; Phage_fiber_1.
CC InterPro; IPR005068; Phage_fiber_2.
CC Pfam; PF03335; Phage_fiber_1.
CC Pfam; PF03406; Phage_fiber_2; 1.
CC KW Fiber protein.
SQ SEQUENCE 774 AA; 77527 MW; CDDIDF85E919123B CRC64;

Query Match 15.8%; Score 74; DB 1; Length 774;
Best Local Similarity 32.2%; Pred. No. 6.8;
Matches 29; Conservative 11; Mismatches 40; Indels 10; Gaps 2;

QY 6 AHIPQLV--ASQSAFAAKAGLMRHRTIGQAEQAAMSAQAFHQGESSAAFOAHARFVAAAA 63
DB 139 AQVALVTDTDSARAAT-----SAGQAASSQAEASSGAEASAKATEAKSAAAA 190

QY 64 KVTLLDVAQNLGEAAGTVVADAAAAT 93
DB 191 ESSKNAATSGAAGAKTSETNAAASQQAAT 220

RESULT 12
BRCA_DROME STANDARD; PRT; 880 AA.
AC Q24206; O40604; Q9W575;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Broad-complex core-protein isoform 6.
GN BR OR BR-C OR EG:17A9.1 OR EG:25D2.1 OR EG:123Fil.1 OR
GN CG11491/CG11514.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., DEVELOPMENTAL STAGE, AND CHARACTERIZATION OF
RP ISOFORMS.
RC TISSUE=Imaginal disks, and Larva;
RX MEDLINE=96299417; PubMed=8660872;

```

```

RA Bayer C.A., Holley B., Fristrom J.W.;
RT "A switch in broad-complex zinc-finger isoform expression is regulated
RT posttranscriptionally during the metamorphosis of Drosophila imaginal
RT discs.";
RL Dev. Biol. 177:1-14(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Oregon-R;
RX MEDLINE=20196011; PubMed=10731137;
RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D., Cadieu E.,
RA Bartell B.G., Ferraz C., Vidal S., Brun C., Denaillies J., Borkova D.,
RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablo S.,
RA Modolell J., Peter A., Schoettler P., Werner M., Mourikioi F.,
RA Beiernt N., Dowe G., Schaefer U., Jaesckle H., Bucheton A.,
RA Callister D.J., Campbell L.A., Darlamitsou A., Henderson N.S.,
RA McMillan P.J., Salles C., Tait E.A., Valentini P., Saunders R.D.C.,
RA Glover D.M.;
RT "From sequence to chromosome: the tip of the X chromosome of D.
RT melanogaster.";
RL Science 287:2220-2222(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adamatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin L.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibagwan C.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusse D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [4]
RP CHARACTERIZATION OF ISOFORMS, AND MUTATIONAL ANALYSIS.
RX MEDLINE=97384928; PubMed=9242423;
RA Bayer C.A., von Kalm L., Fristrom J.W.;
RT "Relationships between protein isoforms and genetic functions
RT demonstrate functional redundancy at the Broad-Complex during
RT Drosophila metamorphosis.";
RL Dev. Biol. 187:267-282(1997).
CC -!- FUNCTION: BROAD-COMPLEX PROTEINS ARE REQUIRED FOR PUFFING AND

```

TRANSCRIPTION OF SALIVARY GLAND LATE GENES DURING METAMORPHOSIS.
-!- SUBCELLULAR LOCATION: Nuclear.
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=6;
Name=6; Synonyms=BCORE-Z4;
IsoId=Q24206-1; Sequence=Displayed;
Name=1; Synonyms=BCORE-TNT1-Q1-Z1;
IsoId=Q01295-1; Sequence=External;
Name=2; Synonyms=BCORE-Q1-Z1;
IsoId=Q01295-2; Sequence=External;
Name=3; Synonyms=BCORE-Q2-Z1;
IsoId=Q01295-3; Sequence=External;
Name=4; Synonyms=BCORE-Z2;
IsoId=Q01295-4; Sequence=External;
Name=5; Synonyms=BCORE-NS-Z3;
IsoId=Q01295-5; Sequence=External;
-!- DEVELOPMENTAL STAGE: ACCUMULATES TO A HIGH LEVEL AT THE BEGINNING OF THE ECDYSONE RESPONSE, DURING THE METAMORPHOSIS OF IMAGINAL DISKS IN PUFF STAGE 1, AND ABRUPTLY DISAPPEARS AFTER SEVERAL HOURS.
-!- INDUCTION: INDUCED AS A PRIMARY RESPONSE TO 20-HYDROXYECDYSONE IN THIRD INSTAR LARVAL IMAGINAL DISKS.
-!- SIMILARITY: Contains 1 BTB/POZ domain.
-!- SIMILARITY: Contains 2 C2H2-type zinc fingers.
-!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534 TO 619 AND 656 TO 694 DUE TO FRAMESHIFTS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U51585; AAB09760.1; ALT_FRAME.
EMBL; AL009146; CAA15627.1; -.
EMBL; AB003421; AAF45647.1; -.
DR FLYBase; FBgn0000210; br.
DR GO; GO:0006914; P:autophagy; IMP.
DR GO; GO:0007459; P:photoreceptor fate commitment (sensu Drosoph. . .; IMP.
DR GO; GO:0007458; P:progression of morphogenetic furrow (sensu . . .; IMP.
DR GO; GO:0009613; P:response to pest/pathogen/parasite; IMP.
DR InterPro; IPR000210; BTB_POZ.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00651; BTB; 1.
DR Pfam; PF00096; zf-C2H2; 2.
DR SMART; SM00225; BTB; 1.
DR SMART; SM00355; Znf_C2H2; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 2.
DR PROSITE; PS50097; BTB; 1.
KW Nuclear protein; DNA-binding; Developmental protein;
KW Zinc-finger; Metal-binding; Alternative splicing.
FT DOMAIN 32 97
FT ZN_FING 710 733
FT ZN_FING 740 763
FT DOMAIN 203 207
FT DOMAIN 265 268
FT DOMAIN 458 466
FT DOMAIN 584 589
FT DOMAIN 618 621
FT DOMAIN 738 803
FT DOMAIN 821 833
FT DOMAIN 862 867
FT DOMAIN 436 436
FT CONFLICT 621 621
FT CONFLICT 624 624
FT CONFLICT 661 662
FT CONFLICT 678 678
FT CONFLICT 722 722
SEQUENCE 880 AA; 500C0A4A38663A AF CRC64;

Query Match 15.8%; Score 74; DB 1; Length 880;
Best Local Similarity 36.4%; Pred. No. 7.7;
Matches 24; Conservative 6; Mismatches 26; Indels 10; Gaps 2;
QY 27 HTIQAEQAAMSAQAFPHQGESAAFOAHAFVAAAANKVTLTDVAQANLGEAGTYVAA 86
Db 775 HGAG-VSQAGSPGSLHQSLSLSSLS-----AAAAAANNVNVGGSGVGAGNAVAA 824
QY 87 DAAAS 92
Db 825 AAAAA 930
RESULT 13
YJ83 MYCTU STANDARD; PRT; 558 AA.
ID YJ83 MYCTU STANDARD; PRT; 558 AA.
AC Q10873;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PE-PGRS family protein RV1983/MT2036.
GN RV1983 OR MT2036 OR MTCY39.36C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OC NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekai F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajadream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; Z74025; CAA98400.1; -.
EMBL; A5007056; AAK46311.1; -.
PIR; E70756; E70756.
TIGR; MT2036; -.
DR Tuberculist; RV1983; -.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.

```
KW Hypothetical protein; Transmembrane; Complete proteome.
FT DOMAIN 27 53 ALA-RICH.
FT TRANSMEM 332 352 POTENTIAL.
FT TRANSMEM 360 380 POTENTIAL.
SQ SEQUENCE 558 AA; 53738 MW; 17ECBE43778E021E CRC64;

Query Match 15.7%; Score 73.5; DB 1; Length 558;
Best Local Similarity 30.0%; Pred.No.5.6; Indels 5; Gaps 2;
Matches 27; Conservative 12; Mismatches 46; Indels 5; Gaps 2;

QY 8 IPOLVASQSAFAKAGLMRHTIQAFQQAAMSA-QAFHOGESSAAFOAHAF-VAQA 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7 VPEFLTAAADVENIGSTLRAANAAASTTALAAGADEVSAVAALPARGQGVQAVS 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 63 AKVNTLLDVAQANLGEAAGTYVAADAAS 92
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 AGAFAHQFQVQTLNLSAGSGYAAAEATIAS 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
RF1_MYCTU STANDARD; PRT; 357 AA.
AC Q10603;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Peptide chain release factor 1 (RF-1).
GN PRFA OR RV1299 OR MT1338 OR WTCY373.19 OR MB1331.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Allard D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -1- FUNCTION: Peptide chain release factor 1 directs the termination
CC of translation in response to the peptide chain termination codons
CC UAG and UAA (By similarity).
```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2004, 17:09:52 ; Search time 39 Seconds
(without alignments)
784.750 Million cell updates/sec

Title: US-09-886-349A-24

Perfect score: 468
Sequence: 1 MSLLDAHIPQLVASOSAFAA.....EAAGTYVAADAAASTYTGTF 97

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPREMBL 25:**

1: sp.archaea.*
2: sp.bacteria.*
3: sp.fungi.*
4: sp.human.*
5: sp.invertebrate.*
6: sp.mammal.*
7: sp.mhc.*
8: sp.organelle.*
9: sp.phage.*
10: sp.plant.*
11: sp.rodent.*
12: sp.virus.*
13: sp.vertebrate.*
14: sp.unclassified.*
15: sp.virus.*
16: sp.bacteriaph.*
17: sp.archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	468	100.0	97	16	O53692 mycobacteri
2	468	100.0	97	16	Q7U2C8 mycobacteri
3	442	94.4	97	16	O53267 mycobacteri
4	442	94.4	97	16	Q7TXXF mycobacteri
5	442	94.4	102	16	Q8VJ87 mycobacteri
6	356	76.1	98	16	Q9CD32 mycobacteri
7	92	19.7	135	16	P95012 mycobacteri
8	92	19.7	135	16	Q7TVC5 mycobacteri
9	92	19.7	373	16	Q8VJG4 mycobacteri
10	89	19.0	457	16	Q7U0X5 mycobacteri
11	87	18.6	1156	16	Q9Z5A4 streptomyc
12	86	18.4	331	16	O53891 mycobacteri
13	86	18.4	335	16	Q8VKA6 mycobacteri
14	86	18.4	476	16	O06199 mycobacteri
15	85	18.2	1147	16	O87848 streptomyc
16	84	17.9	335	16	Q7U0X7 mycobacteri

17	84	17.9	457	16	O53693
18	84	17.9	464	16	Q7TY83 mycobacteri
19	84	17.9	476	16	Q8VKA5 mycobacteri
20	80.5	17.2	382	16	P71748 mycobacteri
21	80.5	17.2	394	16	Q7TXX5 mycobacteri
22	80.5	17.2	462	16	O33310 mycobacteri
23	79.5	17.0	92	13	Q9DF23 myoxocephal
24	79.5	17.0	3703	5	Q8IMW2 Q8IMW2 drosophila
25	79.5	17.0	3912	5	Q9VCS6 Q9VCS6 drosophila
26	79	16.9	113	16	Q8PMS9 Q8PMS9 xanthomonas
27	79	16.9	113	16	Q8PB30 Q8PB30 xanthomonas
28	77.5	16.6	393	16	Q7TJ23 Q7TJ23 mycobacteri
29	77.5	16.6	394	16	O33205 Q33205 mycobacteri
30	77.5	16.6	394	16	Q7TZR6 Q7TZR6 mycobacteri
31	77.5	16.6	411	16	O53939 Q53939 mycobacteri
32	77.5	16.6	956	16	Q8PGH7 Q8PGH7 xanthomonas
33	77	16.5	459	16	Q88819 Q88819 pseudomonas
34	77	16.5	753	2	Q845W1 Q845W1 burkholderi
35	77	16.5	951	5	Q9VEC6 Q9VEC6 drosophila
36	77	16.5	984	5	Q8IN89 Q8IN89 drosophila
37	77	16.5	1366	16	Q8Y120 Q8Y120 raietonia s
38	76.5	16.3	107	10	Q9FEY2 Q9FEY2 heterocapsa
39	76.5	16.3	172	16	Q8VKF1 Q8VKF1 mycobacteri
40	76.5	16.3	172	16	Q7UID8 Q7UID8 mycobacteri
41	76.5	16.3	175	16	O53805 O53805 mycobacteri
42	76.5	16.3	201	10	Q9FEY3 Q9FEY3 heterocapsa
43	76.5	16.3	354	2	Q8AEZ5 Q8AEZ5 frankia sp.
44	76.5	16.3	790	16	Q8ZQF5 Q8ZQF5 salmonella
45	76	16.2	376	16	Q8Z8C1 Q8Z8C1 salmonella

ALIGNMENTS

RESULT 1

ID	O53692	PRELIMINARY;	PRT;	97 AA.
AC	O53692;			
DT	01-JUN-1998 (TRENBLrel. 06, Created)			
DT	01-JUN-1998 (TRENBLrel. 06, Last sequence update)			
DT	01-OCT-2003 (TRENBLrel. 25, Last annotation update)			
DE	Hypothetical protein (PE family protein).			
GN	RV0287 OR MTW035.15 OR MT0300.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCHI_taxid=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RX	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,			
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Suiston J.E., Taylor K., Whitehead S., Barrell B.G.,			
RI	"Deciphering the biology of Mycobacterium tuberculosis from the			
RI	complete genome sequence."			
RL	Nature 393:537-544(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,			
RA	Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W.,			
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains"			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			

```

DR  EMBL; AL021930; CAA17362.1; -.
DR  EMBL; AE006937; AAK44524.1; -.
DR  PIR; E70836; E70836.
DR  TIGR; MT0300; -.
DR  KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 97 AA; 9778 MW; 927527DA610A1637 CRC64;

  Query Match      100.0%; Score 468; DB 16; Length 97;
  Best Local Similarity 100.0%; Pred. No. 3.3e-37;
  Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAFTQGSSAAFAQAAHARFVA 60
DB  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAFTQGSSAAFAQAAHARFVA 60

QY  61 AAQKVTLLDVAQNLGEAAGTYVAADAAAASYTGF 97
DB  61 AAQKVTLLDVAQNLGEAAGTYVAADAAAASYTGF 97

RESULT 2
QYU2C8      PRELIMINARY;      PRT;      97 AA.
AC  QYU2C8;
DT  01-OCT-2003 (TrEMBLrel. 25, Created)
DT  01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DB  Conserved hypothetical protein TB9.8.
GN  ESXG OR YB0255
OS  Mycobacterium bovis.
OC  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC  Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX  NCBI_TaxID=1765;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=AF2122/97;
RX  MEDLINE=22709107; PubMed=12788972;
RA  Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
RA  Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA  Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA  Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA  "The complete genome sequence of Mycobacterium bovis.";
RT  Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR  EMBL; BX248335; CAD93159.1; -.
DR  KW  Complete proteome.
SQ  SEQUENCE 97 AA; 9778 MW; 927527DA610A1637 CRC64;

  Query Match      100.0%; Score 468; DB 16; Length 97;
  Best Local Similarity 100.0%; Pred. No. 3.3e-37;
  Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAFTQGSSAAFAQAAHARFVA 60
DB  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAFTQGSSAAFAQAAHARFVA 60

QY  61 AAQKVTLLDVAQNLGEAAGTYVAADAAAASYTGF 97
DB  61 AAQKVTLLDVAQNLGEAAGTYVAADAAAASYTGF 97

RESULT 3
OS3267      PRELIMINARY;      PRT;      97 AA.
AC  OS3267;
DT  01-JUN-1998 (TrEMBLrel. 06, Created)
DT  01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT  01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DB  PE-family protein.
GN  RV3020C OR MTV012.34C.
OS  Mycobacterium tuberculosis.
OC  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC  Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX  NCBI_TaxID=1765;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=AF2122/97;
RX  MEDLINE=22709107; PubMed=12788972;
RA  Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
RA  Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA  Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA  Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA  "The complete genome sequence of Mycobacterium bovis.";
RT  Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR  EMBL; BX248335; CAD93159.1; -.
DR  KW  Complete proteome.
SQ  SEQUENCE 97 AA; 9778 MW; 927527DA610A1637 CRC64;

  Query Match      100.0%; Score 468; DB 16; Length 97;
  Best Local Similarity 100.0%; Pred. No. 3.3e-37;
  Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAFTQGSSAAFAQAAHARFVA 60
DB  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAFTQGSSAAFAQAAHARFVA 60

QY  61 AAQKVTLLDVAQNLGEAAGTYVAADAAAASYTGF 97
DB  61 AAQKVTLLDVAQNLGEAAGTYVAADAAAASYTGF 97

RESULT 4
QYU2C8      PRELIMINARY;      PRT;      97 AA.
AC  QYU2C8;
DT  01-OCT-2003 (TrEMBLrel. 25, Created)
DT  01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DB  PE family protein.
GN  ESXG OR YB0346C.
OS  Mycobacterium bovis.
OC  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC  Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX  NCBI_TaxID=1765;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=AF2122/97;
RX  MEDLINE=22709107; PubMed=12788972;
RA  Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
RA  Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA  Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA  Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA  "The complete genome sequence of Mycobacterium bovis.";
RT  Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR  EMBL; BX248344; CAD96733.1; -.
DR  KW  Complete proteome.
SQ  SEQUENCE 97 AA; 9842 MW; BA9BCB3180EC17F2 CRC64;

  Query Match      94.4%; Score 442; DB 16; Length 97;
  Best Local Similarity 91.8%; Pred. No. 9.8e-35;
  Matches 89; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAFTQGSSAAFAQAAHARFVA 60
DB  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAFTQGSSAAFAQAAHARFVA 60

QY  61 AAQKVTLLDVAQNLGEAAGTYVAADAAAASYTGF 97
DB  61 AAQKVTLLDVAQNLGEAAGTYVAADAAAASYTGF 97

RESULT 5
QYU2C8      PRELIMINARY;      PRT;      97 AA.
AC  QYU2C8;
DT  01-OCT-2003 (TrEMBLrel. 25, Created)
DT  01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DB  PE family protein.
GN  ESXG OR YB0346C.
OS  Mycobacterium bovis.
OC  Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC  Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX  NCBI_TaxID=1765;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  STRAIN=AF2122/97;
RX  MEDLINE=22709107; PubMed=12788972;
RA  Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
RA  Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA  Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA  Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA  "The complete genome sequence of Mycobacterium bovis.";
RT  Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR  EMBL; BX248344; CAD96733.1; -.
DR  KW  Complete proteome.
SQ  SEQUENCE 97 AA; 9842 MW; BA9BCB3180EC17F2 CRC64;

  Query Match      94.4%; Score 442; DB 16; Length 97;
  Best Local Similarity 91.8%; Pred. No. 9.8e-35;
  Matches 89; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAFTQGSSAAFAQAAHARFVA 60
DB  1 MSLDDAHIPOLVASQSFAFAKAGLMRHTIGQAEQAAMSAQAFTQGSSAAFAQAAHARFVA 60

QY  61 AAQKVTLLDVAQNLGEAAGTYVAADAAAASYTGF 97
DB  61 AAQKVTLLDVAQNLGEAAGTYVAADAAAASYTGF 97

```



```

RESULT 5
Q8VJ87
ID Q8VJ87 PRELIMINARY; PRT; 102 AA.
AC Q8VJ87;
DT 01-NAR-2002 (TrEMBLrel. 20, Created)
DT 01-NAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PE family protein.
GN MT3105.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1773;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007129; AAK47434.1; -.
DR TIGR; MT3105; -.
SQ SEQUENCE 102 AA; 10372 MW; 03CB984D6633D2E4 CRC64;

Query Match 94.4%; Score 442; DB 16; Length 102;
Best Local Similarity 91.8%; Pred. No. 1e-34; 3; Indels 0; Gaps 0;
Matches 89; Conservative

QY 1 MSLLDAHIPOLVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAFAQAAHARFVA 60
DB 6 MSLLDAHIPOLVASHAFPAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAFAQAAHARFVA 65

QY 61 AAKVNTLLDVAQANLGEAGTYVAADAAASTYTG 97
DB 66 AAKVNTLLDVAQANLGEAGTYVAADAAASTYTG 102

RESULT 6
Q9CD32
ID Q9CD32 PRELIMINARY; PRT; 98 AA.
AC Q9CD32;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PE-family protein.
GN M2532.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1769;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=IN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Halroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrall B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011 (2001).
DR EMBL; AL583926; CAC32063.1; -.
DR FIR; A87226; A87226.

Query Match 19.7%; Score 92; DB 16; Length 135;
Best Local Similarity 36.6%; Pred. No. 0.29;
Matches 30; Conservative 8; Mismatches 28; Indels 16; Gaps 2;

QY 11 LVASQSAFAAKAGLMRHTTIGQAEQAAMSAQAFHQGESSAAFAQAAHARFVAANKVNTLLD 70
DB 59 LAGLSAAVAPSG-----AVLSWQA-----NAVAVNAAHARAGAAAAVSAEMR 102

QY 71 VAQANLGEAGTYVAADAAAS 92
DB 103 ATAAALGOARRYAGQDTAAAA 124

RESULT 8
Q7TYC5
ID Q7TYC5 PRELIMINARY; PRT; 135 AA.
AC Q7TYC5;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical alanine rich protein.
GN MB2370.
```

OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RA "The complete genome sequence of Mycobacterium bovis";
 RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 RL EMBL; BX248343; CAD94755.1; -;
 DR Complete proteome.
 KW SEQUENCE 135 AA; EEDC03B2B70DBCC CRC64;

Query Match 19.7%; Score 92; DB 16; Length 135;
 Best Local Similarity 36.6%; Pred. No. 0.29;
 Matches 30; Conservative 8; Mismatches 28; Indels 16; Gaps 2;
 QY 11 LVASQSAFAKAGLMRHHTIGQAQAAQAFHQESSAAFOAHARFVAAAANKVNTLLD 70
 Db 59 LAGELSAVAPSG-----AVLSWQA-----NAVAVNAHARAGAAAASVARNR 102
 QY 71 VAQNLGEAAGTYVAADAAAAS 92
 Db 103 ATAAALGQAARYAGQDTAAAA 124

RESULT 9
 Q8VUG4 PRELIMINARY; PRT; 373 AA.
 AC Q8VUG4;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein MT2615.1.
 GN MT2615.1.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gail J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE007097; AAK46926.1; -;
 DR TIGR; MT2615; -;
 KW Hypothetical protein.
 SQ SEQUENCE 373 AA; BFA60E4609F2E908 CRC64;

Query Match 19.7%; Score 92; DB 16; Length 373;
 Best Local Similarity 36.6%; Pred. No. 0.93;
 Matches 30; Conservative 8; Mismatches 28; Indels 16; Gaps 2;
 QY 11 LVASQSAFAKAGLMRHHTIGQAQAAQAFHQESSAAFOAHARFVAAAANKVNTLLD 70
 Db 59 LAGELSAVAPSG-----AVLSWQA-----NAVAVNAHARAGAAAASVARNR 102
 QY 71 VAQNLGEAAGTYVAADAAAAS 92
 Db 103 ATAAALGQAARYAGQDTAAAA 124

RESULT 10
 Q7UOX5 PRELIMINARY; PRT; 457 AA.
 AC Q7UOX5;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE PE-FGRS: family protein.
 GN PE_FGRS18 OR MB1006C.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RA "The complete genome sequence of Mycobacterium bovis";
 RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 RL EMBL; BX248337; CAD93867.1; -;
 DR Complete proteome.
 KW SEQUENCE 457 AA; 42096 MW; 429E35E2F624BA56 CRC64;
 SQ SEQUENCE 457 AA; 42096 MW; 429E35E2F624BA56 CRC64;
 Query Match 19.0%; Score 89; DB 16; Length 457;
 Best Local Similarity 33.0%; Pred. No. 2.3;
 Matches 29; Conservative 16; Mismatches 39; Indels 4; Gaps 2;
 QY 9 POLVASQSAFAKAGLMRHHTIGQAQAAQAFHQESSAAFOA---AHAR-FVAAAANK 64
 Db 8 POLVSTAADAARIGSINTANTAAATTVLAAAQDEVSTAAALFGSHGQHYQAISAQ 67
 QY 65 VNTLLDVAQANLGEAAGTYVAADAAAAS 92
 Db 68 VAAVQERFVLALSQASSTYAVAAASAT 95
 RESULT 11
 Q9Z5A4 PRELIMINARY; PRT; 1156 AA.
 AC Q9Z5A4;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative secreted protein.
 GN SC06198 OR SC2G5.19.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Oliver K., Harris D.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmid and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]

```
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -!- COFACTOR: ZINC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
DR EMBL; AL939126; CAB36606.1; --.
DR PIR; T34852; T34852.
DR GO; GO:0004024; F:alcohol dehydrogenase activity, zinc-dependent; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR002328; ADH zinc.
DR InterPro; IPR005506; DUF312.
DR Pfam; PF03752; ALF; 8.
DR PROSITE; PS00559; ADH_ZINC; 1.
DR Oxidoreductase; Zinc; Complete proteome.
KW OXIDOREDUCTASE; ZINC; 116717 MW; 70469814FB9517F8 CRC64;
SQ SEQUENCE 1156 AA; 116717 MW; 11566; Length 1156;

Query Match 18.6%; Score 87; DB 16; Length 1156;
Best Local Similarity 37.6%; Pred. No. 10;
Matches 35; Conservative 12; Mismatches 34; Indels 12; Gaps 4;

QY 11 LVASQSA--FAAKAGL-----NRHTIGQAEQAAMSAQAAPHQGESSAAAF-QAAHARFVAAA 62
DB 395 LAAATQATQAAAAAGISAAATARDASAAQAQAAVAQAASGAQAQAQAARAAABADAQA 454

QY 63 AK-----VNTLLDVAQANLGEAAGTYVAADAAA 91
DB 455 ARATKAANRAQSLANTAAASAAAARKAADSA 487

RESULT 12
OS3891 PRELIMINARY; PRT; 331 AA.
AC OS3891;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PGSR-family protein.
GN RV0978C OR MT0044.06C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moulé S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL; AL021999; CAAL1577.1; --.
DR PIR; F70820; F70820.
```

```
DR Tuberculist; RV0978C; --.
DR InterPro; IPR001258; NHL.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF01436; NHL; 3.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
KW COMPLETE PROTEOME.
SQ SEQUENCE 331 AA; 31019 MW; A093EC12F1691CA5 CRC64;

Query Match 18.4%; Score 86; DB 16; Length 331;
Best Local Similarity 33.0%; Pred. No. 3;
Matches 29; Conservative 14; Mismatches 41; Indels 4; Gaps 2;

QY 9 POLVASQSAFAAKAGLMRHTIGQAEQAAMSAQAAPHQGESSAAAFQA---AHAR-FVAAA 64
DB 8 POLVSTAAADAARIGSAINTANTAAATTQVLAQAQDEVSTAALFGSHGQHYQAISAQ 67

QY 65 VNTLLDVAQANLGEAAGTYVAADAAAAS 92
DB 68 VAAVQQRFLVLSQAGSTYAVAEASAT 95

RESULT 13
OS3891 PRELIMINARY; PRT; 335 AA.
AC OS3891;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PE_PGSR family protein.
GN MT1006.1.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Deicher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE006985; AAK45254.1; --.
DR TIGR; MT1006; --.
DR InterPro; IPR001258; NHL.
DR InterPro; IPR000084; PE_region.
DR Pfam; PF01436; NHL; 3.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; PE_region; 1.
DR ProDom; PD001223; PE_region; 1.
SQ SEQUENCE 335 AA; 31351 MW; C2C742369C499244 CRC64;

Query Match 18.4%; Score 86; DB 16; Length 335;
Best Local Similarity 33.0%; Pred. No. 3.1;
Matches 29; Conservative 14; Mismatches 41; Indels 4; Gaps 2;

QY 9 POLVASQSAFAAKAGLMRHTIGQAEQAAMSAQAAPHQGESSAAAFQA---AHAR-FVAAA 64
DB 8 POLVSTAAADAARIGSAINTANTAAATTQVLAQAQDEVSTAALFGSHGQHYQAISAQ 67

QY 65 VNTLLDVAQANLGEAAGTYVAADAAAAS 92
DB 68 VAAVQQRFLVLSQAGSTYAVAEASAT 95

RESULT 14
OS3891 PRELIMINARY; PRT; 476 AA.
AC OS3891;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
```

